Review of Economic Studies (2022) **89**, 1420–1459 doi:10.1093/restud/rdab061

Advance access publication 23 September 2021

Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data

BENJAMIN FABER
Department of Economics, UC Berkeley and NBER

and

THIBAULT FALLY
Department of Agricultural and Resource Economics, UC Berkeley and NBER

First version received July 2017*; Editorial decision July* 2020*; Accepted July* 2020 (*Eds.*)

A growing literature has documented the role of firm heterogeneity within sectors for nominal income inequality. This article explores the implications for household price indices across the income distribution. Using detailed matched U.S. home and store scanner microdata, we present evidence that rich and poor households source their consumption differently across the firm size distribution within disaggregated product groups. We use the data to examine alternative explanations, propose a tractable quantitative model with two-sided heterogeneity that rationalizes the observed moments, and calibrate it to explore general equilibrium counterfactuals. We find that larger, more productive firms sort into catering to the taste of richer households, and that this gives rise to asymmetric effects on household price indices. We quantify these effects in the context of policy counterfactuals that affect the distribution of disposable incomes on the demand side or profits across firms on the supply side.

Key words: Firm heterogeneity, Real income inequality, Household price indices, Scanner data.

JEL Codes: F15, F61, E31

1. INTRODUCTION

Income inequality has been on the rise in the U.S. and many other countries, attracting the sustained attention of policy makers and the general public. In this context, a growing literature has documented the role of Melitz-type firm heterogeneity within sectors in accounting for nominal income inequality.[1] In this article, we explore the implications of firm heterogeneity for household price indices across the income distribution. We aim to answer three central questions: (1) to

1. For example, Song *et al.* (2018), Helpman *et al.* (2017), and Card *et al.* (2013). See related literature at the end of this section.

The editor in charge of this paper was Adam Szeidl.

what extent do rich and poor households source their consumption baskets from different parts of the firm size distribution?; (2) what forces explain these differences?; and (3) what are the implications of the answers to (1) and (2) for the impact of policies or economic shocks on real income inequality?

In answering these questions, the article makes three main contributions. First, using detailed matched home and store scanner consumption microdata, we document significant differences in the weighted average firm sizes that rich and poor U.S. households source their consumption from, and explore alternative explanations.[2] Second, to rationalize these moments we develop a tractable quantitative model of product quality choice with two-sided heterogeneity across firms and households, and use the microdata to calibrate it and quantify the forces underlying the stylized facts. Third, we explore general equilibrium (GE) policy counterfactuals to illustrate how, in a setting where rich and poor households source their consumption from heterogeneous firms, policies, and economic shocks give rise to asymmetric price index effects across the income distribution.

At the centre of the analysis lies a detailed collection of microdata that allow us to trace the firm size distribution into the consumption baskets of households across the income distribution. We combine a dataset of 345 million consumer transactions when aggregated to the household-by-retailer-by-barcode-by-half-year level from the Nielsen U.S. Home Scanner (Nielsen Company, 2016a) data over the period 2006–2014, with a dataset of 12.2 billion store transactions when aggregated to the store-by-barcode-by-half-year level from the Nielsen U.S. Retail Scanner data (Nielsen Company, 2016b) covering the same period. The combination of home and store-level scanner data allows us to trace the size distribution of producers of brands—measured in terms of national sales that we aggregate across on average 27,000 retail establishments each half year in the store scanner data—into the consumption baskets of on average 59,000 individual households per half year in the home scanner data within more than 1000 disaggregated retail product modules, such as carbonated drinks, shampoos, pain killers, desktop printers, or microwaves.

The analysis proceeds in four steps. In Step 1, we use the data to document a set of stylized facts. We find that the richest 20 (respectively 10) % of U.S. households source their consumption from on average 20 (respectively 27) % larger producers of brands within disaggregated product groups compared to the poorest 20 (respectively 10) % of U.S. households. We also document that these differences in weighted-average firm sizes arise in a setting where the rank order of household budget shares on different producers within a product group is preserved across income groups—*i.e.* the largest firms command the highest budget shares for all income groups. We interpret these stylized facts as equilibrium outcomes in a setting where both consumers and firms optimally choose their product attributes. We also use the microdata to explore a number of alternative explanations and find that differential coverage of retail consumption or differences in product supply and pricing across the income distribution are unlikely driving these results. We also document that the relationship between incomes and firm sizes holds across different product departments in the data, and explore whether consumption categories not fully represented in the Nielsen data, such as consumer durables, health expenditures, or digital media, show similar patterns. Focusing on subsets of these categories that are covered in the data—household appliances, pharmaceuticals, and video, audio, and software purchases, respectively—we find similar or more pronounced differences in firm sizes across incomes.

In Step 2, we write down a tractable model that rationalizes the stylized facts in the data. On the consumption side, we specify non-homothetic preferences allowing households across the income distribution to differ both in terms of their price elasticities as well as in their evaluations

2. We refer to firm size in terms of relative firm sales within product groups.

of product quality attributes. On the production side, we introduce quality choice into a model of heterogeneous firms within sectors. Both marginal and fixed costs can be functions of output quality, allowing for economies of scale in production. Markups can vary across firms due to both oligopolistic competition and selling to heterogeneous consumers. Firms now operate in a setting where their pricing and quality choices affect the composition of market demand that they face. Modelling product choices with two-sided heterogeneity implies that shocks that affect producers differently, such as trade integration, corporate taxes, and regulations, can feed into the consumption baskets of rich and poor households asymmetrically. In turn, changes in the distribution of disposable incomes, due to *e.g.* income tax reform, affect firms differently across the size distribution with GE knock-on effects on household price indices.

In Step 3, we use the microdata to estimate the preference and technology parameters. On the demand side, we find that rich and poor households differ both in terms of price elasticities and their valuation of product quality attributes. We find that poorer households have higher price elasticities relative to richer households, but that these differences, while statistically significant, are relatively minor. We also find that, while households on average agree on the ranking of quality evaluations across producers, richer households value higher quality significantly more. On the production side, we estimate that producing higher quality increases both the marginal and the fixed costs of production, giving rise to economies of scale in quality production. To estimate the technology parameters, we use two different estimation strategies. The first follows existing work, and is based on cross-sectional variation in firm scale and output quality. The second exploits within-firm changes in brand quality and scale over time. To identify the effect of firm scale on output quality in the panel estimation, we use state-level measures of changes in brand quality on the left-hand side, and construct an instrument (IV) for national brand scale on the right-hand side. The IV exploits pre-existing differences in the geography of brand sales across other US states interacted with state-level variation in average sales growth observed in other product groups.

The parameter estimates from Step 3 reveal two opposing forces that in equilibrium determine both the sorting of firms across product quality attributes and firm sizes across consumption baskets. On the one hand, larger firms offer lower quality-adjusted prices, which increases the share of their sales coming from more price-sensitive lower-income consumers. Since these consumers value quality relatively less, this channel, ceteris paribus, leads poorer households to source their consumption from on average larger firms, which in turn pushes these firms to produce at lower output quality. On the other hand, the estimated economies of scale in quality production give larger firms incentives to sort into higher product quality, catering to the taste of richer households. Empirically, we find that this second channel dominates the first, giving rise to the endogenous sorting of larger, more productive firms into products that are valued relatively more by richer households. Armed with these estimates, we find that the observed stylized facts from step 1 translate into significant differences in the weighted-average product quality and quality-adjusted prices embodied in consumption baskets across the income distribution. The richest 20% of U.S. households source their consumption from on average 22% higher-quality producers compared to the poorest quintile of households. At the same time, we find that the richest income quintile source their consumption at on average 10% lower quality-adjusted prices. Overall, we find that the calibrated model based on the estimates from Step 3 fits the observed differences in firm sizes across consumption baskets from step 1 both qualitatively and quantitatively.

In the final Step 4, we use the calibrated model to quantify a new set of price index implications in the context of two policy counterfactuals. The first counterfactual affects the distribution of disposable incomes on the demand side and is motivated by the recent debate about progressive income taxation in the U.S. and elsewhere. We evaluate the price index effects of returning from current U.S. tax rates to a more progressive post-WWII U.S. tax schedule, increasing the effective rate on the richest household group in our calibration from currently around 30–50 percent. This policy also closely corresponds to the counterfactual of moving the U.S. to the current average effective rate on this group among Northern European countries, and it is in line with the proposed tax reforms of two presidential candidates for the 2020 U.S. elections (Sanders and Warren). We find that the resulting compression of disposable incomes gives rise to GE knock-on effects—through changes in firm scale, output quality, variable markups, and exit/entry across the firm size distribution—that affect rich and poor households differently, amplifying the progressivity of the reform. As a result, the poorest 20% of U.S. households experience a 3 percentage point lower inflation rate for retail consumption compared the richest 20%.[3]

The profits of firms across the size distribution on the supply side. The counterfactual is motivated by the ongoing debate about closing loopholes in corporate taxation. A growing literature in public finance has documented larger possibilities for tax avoidance or evasion among large U.S. corporations (*e.g.* Bao and Romeo, 2013; Guvenen *et al.*, 2017; Wright and Zucman, 2018). We use these findings to evaluate the implications of eliminating the kink that has been documented at the 95th percentile of the firm size distribution in the otherwise smoothly increasing relationship between firm size and effective corporate tax rates. This policy would lead to an increase of on average 5% in corporate taxes paid by the largest 5% of producers, ranging between on average 1% at the 95th percentile to 11% at the 99th. We find that even such a relatively modest adjustment in corporate taxation leads to a meaningful GE effect on inflation differences between rich and poor households. This is in the order of a 1.5 percentage point lower cost of living increase for retail consumption among the bottom 20% of U.S. households compared to the top 20%. We document that the direct incidence of this policy—holding initial household and firm decisions fixed—accounts for about one third of the inflation difference, with the remainder due to endogenous firm adjustments that affect consumption baskets differently.

We also explore the implications for the distribution of the gains from trade, extending a textbook Melitz model with two symmetric countries to quality choice with two-sided heterogeneity within countries, and calibrating it to the U.S. microdata. We find that allowing for firm heterogeneity across consumption baskets makes the gains from trade more unequal: increasing bilateral import penetration by 10 percentage points leads to a 3.5 percentage point lower retail inflation for the richest 20% of households compared to the poorest 20%. Taken together, these findings illustrate that firm heterogeneity affects inequality in more complex ways than through the nominal earnings of workers that have been the focus of the existing literature. These insights arise after introducing a basic set of features that we observe in the microdata—allowing for product attribute choice by heterogeneous firms and households—into an otherwise standard economic environment.

This paper is related to the growing literature on the extent, causes and consequences of firm heterogeneity within sectors that has spanned different fields in economics, including international trade (Melitz, 2003; Bernard *et al.*, 2007), industrial organization (Bartelsman *et al.*, 2013), macroeconomics (Hsieh and Klenow, 2009), development (Peters, 2020), labour economics (Card *et al.*, 2013), and management (Bloom and Van Reenen, 2007). Within this literature, this paper is most closely related to existing work on the implications of firm heterogeneity for nominal income inequality (Frias *et al.*, 2009; Davis and Harrigan, 2011; Card *et al.*, 2013; Sampson, 2014; Burstein and Vogel, 2017; Helpman *et al.*, 2017; Song *et al.*, 2018).

Our theoretical framework builds on existing work, both on quality choice across the domestic income distribution with non-homothetic preferences on the demand side (Fajgelbaum *et al.*, 2011; Handbury, 2019) and on quality choice across heterogeneous firms on the supply side (*e.g.*

3. The data allow us to calibrate the model at the level of five broad income groups, while both tax changes and taste for quality increase with incomes within these groups. Results are conservative in this respect.

Johnson, 2012; Kugler and Verhoogen, 2012; Feenstra and Romalis, 2014; Bastos *et al.*, 2018; Fieler *et al.*, 2018). While these building blocks have been used separately—*i.e.* heterogeneous firms facing a representative agent in each country or non-homothetic preferences facing homogeneous firms—this paper combines them in a quantitative GE model. This two-sided heterogeneity within countries allows us to rationalize the observed differences in firm sizes across consumption baskets and gives rise to new price index implications across the income distribution in the light of economic shocks or policy changes.[4]

The article is also related to a growing empirical literature using the Nielsen data (Broda and Weinstein, 2010; Handbury and Weinstein, 2014; Hottman *et al.*, 2016; Handbury, 2019). Most of this literature has been based on the home scanner data. More recently, Argente and Lee (2021) and Jaravel (2019) use the home and store scanner data to document that lower-income households have experienced higher inflation over the past decade and beyond. Argente and Lee (2021) relate this finding to a higher possibility for quality-downgrading among higher-income households during the Great Recession, and Jaravel (2019) to more innovation and competition in product groups consumed relatively more by richer households. In this article, we present new empirical evidence suggesting that the widely documented presence of firm heterogeneity within sectors translates asymmetrically into the consumption baskets of rich and poor households, quantify the underlying forces and explore a new set of implications in the context of policy counterfactuals.

The remainder proceeds as follows. Section 2 describes the data. Section 3 documents stylized facts and explores alternative explanations. Section 4 presents the model. Section 5 presents parameter estimation. Section 6 presents the counterfactual analysis. Section 7 concludes.

2. DATA

2.1. *Retail scanner data*

The Nielsen retail scanner data consist of weekly price and quantity information generated by point-of-sale systems for more than 100 participating retail chains across all U.S. markets between January 2006 and December 2014. When a retail chain agrees to share their data, all of their stores enter the database. As a result, the database includes more than 50,000 retail establishments. The stores in the database vary widely in terms of formats and product types: *e.g.* grocery, drug, mass merchandizing, appliances, liquor, or convenience stores.

Data entries can be linked to a store identifier and a chain identifier so a given store can be tracked over time and can be linked to a specific chain. While each chain has a unique identifier, no information is provided that directly links the chain identifier to the name of the chain. This also holds for the home scanner dataset described below. The implication of this is that the product descriptions and barcodes for generic store brands within product modules have been anonymized. However, both numeric barcode and brand identifiers are still uniquely identified, which allows

4. A notable exception is Feenstra and Romalis (2014) who combine quality choice by heterogeneous firms under monopolistic competition with non-homothetic preferences on the demand side. Since their objective is to infer quality from observed unit values in international trade flows, their model features a representative agent within countries on the demand side. On the supply side, they follow *e.g.* Baldwin and Harrigan (2011) and Kugler and Verhoogen (2012) modelling quality choice as a deterministic function of a firm's productivity draw. To be able to rationalize our empirical results on the effect of firm scale on quality upgrading in firm-level panel data, we instead allow fixed costs to be a function of output quality, following earlier work by Sutton (1998) and Hallak and Sivadasan (2013) and the second model variant in Kugler and Verhoogen (2012).

us to observe sales for individual barcodes of generic store brands within each product module in the same way we observe sales for non-generic products.

When aggregated to the store-by-barcode-by-half-year level, each half year covers on average \$113 billion worth of retail sales across 27,000 individual stores in more than 1000 disaggregated product modules, 2500 U.S. counties and across more than 730,000 barcodes belonging to 175,000 producers of brands (see Table A.1).[5] As described in more detail in the following section, we use these data in combination with the home scanner data described below in order to trace the distribution of firm sizes (measured in terms of national sales measured across on average 27,000 stores per half year) into the consumption baskets of individual households.

2.2. *Home scanner data*

The Nielsen home scanner data is collected through hand-held scanner devices that households use at home after their shopping in order to scan each individual transaction they have made. Importantly, the home and store level scanner datasets can be linked: they use the same codes to identify retailers, product modules, product brands as well as barcodes. As described in more detail in the following section, we use this feature of the database to estimate weighted average differences in firm sizes across consumption baskets.

When aggregated to the household-by-barcode-by-half-year level, each half year covers on average \$109 million worth of retail sales across 59,000 individual households in more than 1000 product modules, 2600 US counties and close to 600,000 barcodes belonging to 185,000 producers of brands (Table A.1). One shortcoming of the home scanner dataset is that nominal household incomes are measured imprecisely. First, incomes are reported only across discrete income ranges. And those income bins are measured with a two-year lag relative to the observed shopping transactions in the dataset. To address this issue, we divide households in any given half year into percentiles of total retail expenditure per capita.[6] To address potential concerns about a non-monotonic (or decreasing) relationship between total retail outlays and incomes, we also confirm that our measure of total retail expenditure per capita is monotonically increasing in reported nominal incomes two years prior (confirming existing evidence that retail expenditure has a positive income elasticity) (Figure A.1).

These descriptive statistics also help clarify the relative strengths and weaknesses of the two Nielsen datasets. The strength of the home scanner database is the detailed level of budget share information that it provides alongside household characteristics. Its relative weakness in comparison to the store-level retail scanner data is that the home scanner sample of households only covers a small fraction of the U.S. retail market in any given period. Relative to the home scanner data, the store-level retail scanner data cover more than 1000 times the retail sales in each half year. This article takes advantage of both datasets for the empirical analysis, by combining national sales by product from the store scanner data with the detailed information on individual household consumption shares in the home scanner data.

5. We do not make use of Nielsen's "Magnet" database that covers non-barcoded products, such as fresh produce.

6. Per capita expenditure can be misleading due to non-linearities in per capita outlays with respect to household size (*e.g.* Subramanian and Deaton, 1996). To address this concern, we non-parametrically adjust for household size by first regressing log total expenditure on dummies for each household size with a household size of 1 being the reference category and a full set of household socio-economic controls. We then deflate observed household total expenditure to per capita equivalent expenditure by subtracting the point estimate of the household size dummy (which is non-zero and positive for all households with more than one member).

3. STYLIZED FACTS

This section draws on the combination of the home scanner and retail scanner data to document a set of stylized facts about firm heterogeneity in the consumption baskets of households across the income distribution. Supplementary Appendix figures and tables are prefixed by "A."

3.1. *Definition of firms*

Using both datasets, we can document what has been shown many times in manufacturing establishment microdata (Bernard *et al.*, 2007; Bartelsman *et al.*, 2013): firm sizes (measured here in terms of sales) differ substantially within disaggregated product groups (Figure A.2). We define a firm as a producer of a brand within product modules in the Nielsen data. This leads to an average of about 150 active firms within a given product module. Two possible alternatives would be to define a firm as a barcode product (leading to on average 700 firms per module), or as a holding company (leading to on average less than 40 firms per module). To fix ideas, an example for the product module Shampoo would be the barcode product Ultimate Hydration Shampoo (22 oz bottle) belonging to brand TRESemmé that is owned by the holding company Unilever.

We choose the definition of firms as brands within product modules for two main reasons, and then check the robustness of our findings to alternative definitions. First, our objective is to define a producer within a given module as closely as possible to an establishment in commonly used manufacturing microdata, following the bulk of the recent literature on firm heterogeneity (Melitz, 2003). The definition of firms as holding companies (*e.g.* Procter&Gamble) would be problematic as these conglomerates operate across hundreds of brands produced in different establishments. The definition of firms at the barcode level would be problematic for the opposite reason, because the same establishment produces different pack sizes of the same product that are marked by different barcodes. In this light, defining producers of brands within disaggregated product modules as firms is likely the closest equivalent to observing several different establishments operating in the same disaggregated product group. Second, our theoretical framework features endogenous investments in product quality across firms, and it is at the level of brands within product groups that these decisions appear to be most plausible.[7]

The distribution of national market shares measured using the home scanner data (on average 59,000 households per half year) appears to be compressed relative to that measured using the store scanner data (on average 27,000 thousand stores per half year) (Figure A.2). This compression is stronger before applying the Nielsen household sampling weights, but still clearly visible after applying the weights. It also holds when restricting attention to producers of brands observed in both datasets. In the following, we report the main new stylized fact using the firm size distributions computed from both datasets. Given that the retail scanner data capture more than 1000 times the amount of transactions compared to the home scanner data, we then choose the store scanner data as our preferred measure of brand-level national market shares.

7. To corroborate this, we confirm in the data that 95% of variation in the average unit value paid for barcodes within product modules is accounted by for by brand-by-pack-size fixed effects. Similarly, on average 80% of the sum of absolute differences in expenditure shares between the richest and poorest household quintiles across all barcodes within product modules are accounted for by differences in budget shares across brands (leaving 20% to be explained by differences within brands).

3.2. *Firm heterogeneity across consumptionbaskets*

Figure 1 depicts the main stylized fact of the article. Pooling repeated cross-sections across 18 half-year periods, we depict percentiles of household per capita expenditure on the x-axis and weighted average deviations of log firm sales from the product module-by-half-year means on the y-axis.[8] The weights correspond to each household's retail consumption shares across all brands in all product modules consumed during the six-month period. When collapsed to five per capita expenditure quintiles on the right panel of Figure 1, we find that the richest 20% of U.S. households source their consumption from on average 20% larger producers of brands within disaggregated product modules compared to the poorest 20%. These figures are our preferred measure of the national firm size distribution using the store scanner data. But, as the figure shows, a very similar relationship holds when using the firm size distribution from the home scanner data instead. This relationship is monotonic across the income distribution, and the firm size difference increases to 27% when comparing the richest and poorest 10% of households. As discussed above, Figure 1 is also robust to alternative definitions of firms in the Nielsen data: Figure A.3 shows close to identical results when defining firms instead as holding companies within product groups.[9]

What types of shopping decisions are driving these differences in weighted average firm sizes across the income distribution? In Table A.2, we present the brands with the most positive and most negative differences in consumption shares between rich and poor household quintiles across three popular product modules for each of the eight product departments in our consumption microdata. We also list the difference in their log average unit values (price per physical unit) as well as the difference in their national market shares within that product module. Two features stand out. First, the brand that is most disproportionately consumed by the rich has a higher unit value and a larger market share relative to the brand that is most disproportionately consumed by the poor.[10] Second, looking at the brand names it seems that richer households have a tendency to consume from the leading premium brands whereas the poorest quintile of households have a higher tendency to pick either generic store brands, or cheaper second and third-tier brands in the product group (*e.g.* Tropicana versus generic OJ, Pepsi versus generic Cola, Duracell versus Rayovac, Tide versus Purex, Dove versus Dial, Heinz versus Hunt's).

Finally, we investigate whether these observed differences in product choices are driven by a fundamental disagreement about relative product appeal across rich and poor households. Do we see rich households consuming a large share of their expenditure from the largest producers while poor households spend close to none of their budget on those same producers? Or do households from different income groups agree on their relative evaluations of value-for-money across producers, such that the rank order of their budget shares is preserved across the income distribution? Figure A.6 documents that the latter appears to be the case in the data. The figure depicts—separately for each income group—non-parametric estimates of the relationship between income group-specific deviations in log expenditures across brands within product modules on the y-axis and deviations in log total market sales of those same producers in the store scanner data on the x-axis. The fact that expenditure shares within each income group

8. To avoid measurement error from exiting or entering households in the consumer panel, we restrict attention to households for each half-year period that we observe to make purchases in both the first and the final month of the half year.

9. Figure A.4, we shows an alternative representation plotting sales shares from different quintiles of the income distribution on the y-axis across firm sizes on the x-axis. Figure A.5 shows the relationship in Figure 1 after computing firm sizes in terms of quantities (units sold).

10. The scanner data allow the comparison of prices per unit (unit values) for identically measured units across products within product module (*e.g.* litres of milk, units of microwaves, grams of cereal, etc).

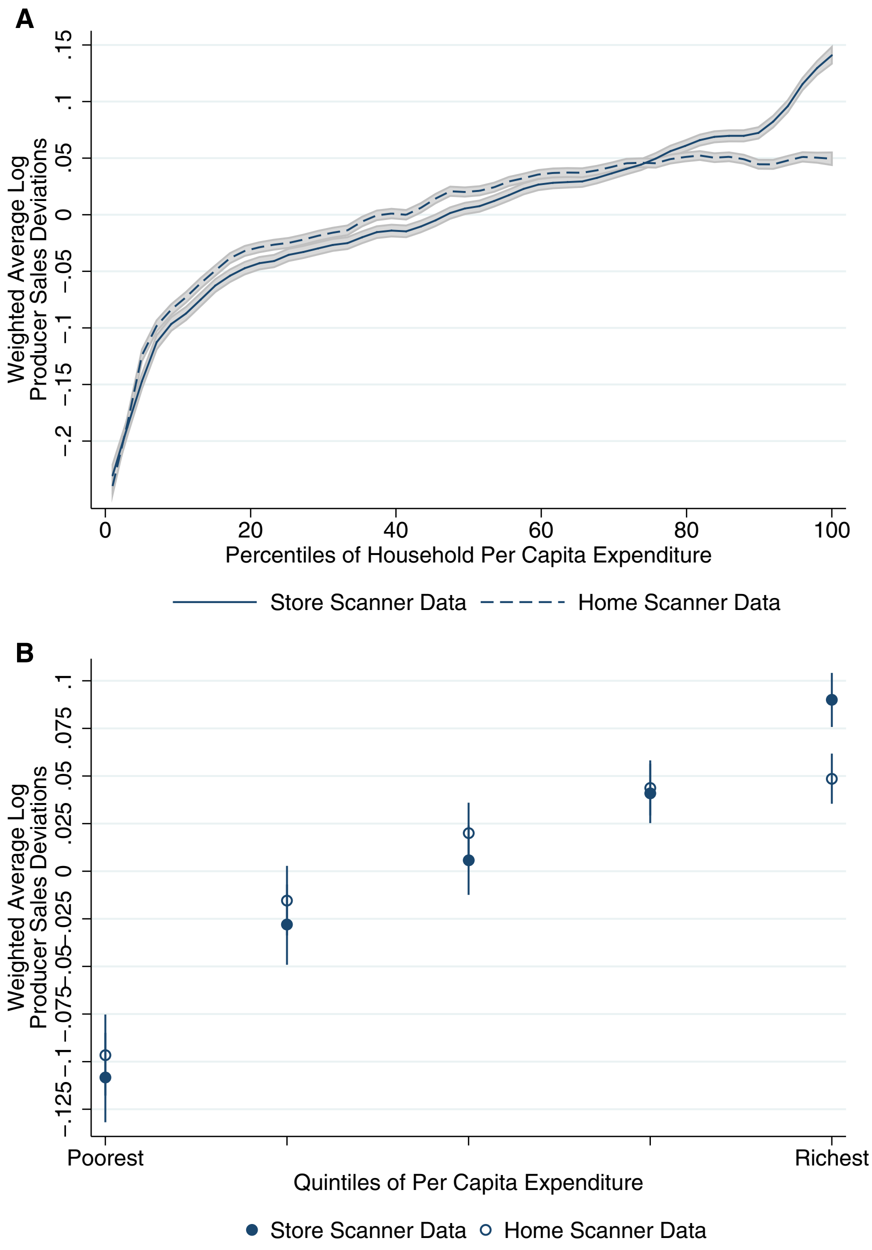


FIGURE 1

Richer households source their consumption from larger firms

Notes: The figure depicts deviations in weighted average log firm sales embodied in the consumption baskets of on average 59,000 US households during 18 half-year periods between 2006 and 2014. The y-axis in both graphs displays weighted average deviations in log producer sales within more than 1000 product modules where the weights are household expenditure shares across producers of brands. In the first step, we compute brand-level deviations from mean log national sales within product module-by-half-year cells from either the home or the store-level scanner data. In the second step, these deviations are then matched to brand-level half-yearly household expenditure weights in the home scanner data. The final step is to collapse these data to weighted average log firm size deviations embodied in household consumption baskets. The x-axis displays national percentiles of per capita total household retail expenditure per half year period (see Section 2). The fitted relationship in Panel A correspond to local polynomial regressions. Panel B displays point estimates across five groups of households. Standard errors in both graphs are clustered at the county level, and the displayed confidence intervals are at the 95% level. Table A.1 provides descriptive statistics. See Section 3 for discussion.

increase monotonically with total firm size suggests that households on average agree on their evaluation of product quality attributes given prices, and that the rank order of budget shares across producers is preserved to a striking extent across all income groups. To express this in a single statistic, we find that the rank order correlation between the richest income quintile and the poorest for rankings of brand market shares within product modules is 0.89 when pooled across all product modules in the data. However, it is also apparent from the difference in slopes depicted in Figure A.6 that while all households spend most of their budget on the largest firms within product modules, richer households spend relatively more of their total budget on these largest producers relative to poorer households.

Related to this, we also investigate the role of the extensive margin in product choice underlying Figure 1. We find that while the average number of both UPCs and brands consumed within a module increase with incomes, the fraction of total retail expenditure accounted for by products consumed across all five income groups is very large (>95%) and close to constant across the income distribution (Figure A.7). Consistent with the rank order correlations above, this suggests that the extensive margin, while visible in the data, is unlikely to play a significant role in accounting for differences in average firm sizes across groups of rich and poor households.[11]

3.3. *Interpretation and alternative explanations*

One natural interpretation of these stylized facts is that they arise as equilibrium outcomes in a setting where both heterogeneous firms and households choose the product attributes they produce or consume. However, there are a number of alternative and somewhat more mechanical explanations that we explore using the microdata before moving on to the model. Another question is how representative these findings are for consumption categories that are not well represented in the Nielsen data.

3.3.1. Representativeness and data-driven explanations. One concern is that the relationship documented in Figure 1 could in part be driven by shortcomings of the data. First, we explore the representativeness of the stylized fact in Figure 1. To this end, we estimate the relationship separately for each of the 18 half-year periods and for each of eight broad product categories in the Nielsen data: beverages, dairy products, dry grocery, frozen foods, general merchandise, non-food grocery, health and beauty, and packaged meat (Figure A.9).[12] We find that the pattern of firm size differences across consumption baskets holds across these different product segments and is not driven by one particular type of consumer products. We also find that the stylized fact in Figure 1 holds in each of the 18 half-year periods in our data.

As reported by *e.g.* Broda and Weinstein (2010) and more recently in Jaravel (2019), the product groups covered in the Nielsen data can be matched to product groups in the US CEX expenditure surveys that account for about 40% of goods consumption, which in turn accounts for about one third of total household outlays. To further explore to what extent the stylized fact in Figure 1 holds among consumption categories not fully represented in the Nielsen data, such as durables consumption, health expenditures or digital media, we use parts of these categories that are covered in the data. In particular, we estimate Figure 1 separately for household appliances (*e.g.*

11. In support of this, Figure A.8 confirms that we find close to identical results after limiting the product space to only products consumed by all income quintiles in a given period (or across all periods).

12. We combine observations for alcoholic and non-alcoholic beverages as one department in these graphs. Our reported findings above hold separately for both of these departments. We pool them here to be consistent with Section 5, where having one combined group for Beverages addresses sparsity in the parameter estimation.

air conditioners, refrigerators, kitchen appliances, desktop printers), pharmaceuticals purchases and audio, video and software purchases (Figure A.10). Reassuringly, we find that the relationship holds similarly or is more pronounced.[13]

Second, it could be the case that generic store brands are produced by the same (large) producers and sold under different labels across retail chains. If poorer households source more of their consumption from generics, then we could under-estimate their weighted average producer size due to this labelling issue. To address this concern, we re-estimate Figure 1 after restricting consumption to sum to 100% for all non-generic product consumption for each household. We find close to identical results compared to the baseline in Figure 1 (Figure A.11).

Third, it could be the case that we are missing systematically different shares of retail consumption across rich and poor households due to the exclusion of products sold by retail chains that are not participating in the store-level retail scanner data (that we use to compute national market sales across producers). For example, it could be the case that richer households purchase a larger share of their retail consumption from independent boutique retailers selling small specialized brands, in which case we would be over-stating the weighted average firm sizes among richer households. Conversely, it could be true that richer households spend on average more on retail chains (compared to *e.g.* corner stores). To address such concerns, we make use of the fact that the home scanner data do not restrict reporting to participating retail chains in the store scanner data.[14] We then re-estimate Figure 1 after only including households for which we observe more than 90, 95, or 97.5% of their total reported retail expenditure on brands that are matched across both data sets, and find that the relationship holds close to identically in all specifications (Figure A.12). To further corroborate, we also show that the fraction of total expenditure on brands in both datasets differs by less than 3% across the income groups (Figure A.13).[15]

Another data-related concern is that the Nielsen data do not allow us to observe firm sales outside the U.S. market. For both U.S.-based exporters and imported brands, we are thus mismeasuring total firm sales relative to domestic-only U.S. producers. Given existing evidence on the selection of firms into trade, it is likely that the resulting measurement error in firm sizes is positively related to the observed U.S. market shares in the Nielsen data (understating true differences in firm sizes). To corroborate this, we report differences in weighted-average firm sizes across incomes separately for product groups both below and above-median import penetration or export shares.[16] We find that the differences in firm sizes across rich and poor households are indeed slightly more pronounced in the below-median sectors for both import or export shares (Figures A.14 and A.15).

3.3.2. Differences in supply and pricing across incomes. Another explanation could be that rich and poor households live in geographically segmented markets and/or shop across segmented store formats, so that differential access to producers, rather than heterogeneous household preferences, could be driving the results. We explore to what extent differences in household geographical location as well as differences in retail formats within locations

13. The Nielsen data unfortunately do not cover services consumption. For future research, newly available credit card microdata could potentially be used to investigate the relationship between card holder incomes and the firm size of service providers they source their consumption from.

14. Since retail chain participation in the store scanner data is not made public by Nielsen, home scanner participants are not made aware of this either.

15. The “missing retailers” concern is also not apparent in Figure 1 that depicts very similar patterns when using 100% of household retail consumption as reported in the home scanner data.

16. To this end, we match the Nielsen product groups to 4-digit SIC codes in 2005 U.S. trade data. See Table A.3. Below median is equivalent to less than 10% for both import penetration and export shares.

play in accounting for Figure 1 (Figure A.16). We first re-estimate the same relationship after conditioning on county-by-half-year fixed effects when plotting the firm size deviations on the y-axis. Second, we additionally condition on individual household consumption shares across 79 different retail store formats (*e.g.* supermarkets, price clubs, convenience stores, pharmacies, liquor stores).[17] We find a very similar relationship compared to Figure 1 in both cases, suggesting that differential access to producers is unlikely to be the driver.

Related to differential access to producers, differential pricing across income groups could be another potential explanation. For this to matter in our context, it would have to be the case that larger and smaller producers differ in their extent of price discrimination across income groups, such that larger firms become relatively more attractive for richer households. To explore this in the data, we plot the relationship between unit values and firm sizes separately for prices paid by the poorest and the richest income quintiles (Figure A.17). If it were true that larger firms offer differentially lower or higher prices compared to smaller firms across high- and low-income groups, we would expect the slope of these relationships to differ between the income groups. Using unit values and firm sizes defined either at the UPC or at the brand-level, we find that this relationship is close to identical for both rich and poor consumers, providing some reassurance that richer households do not pick larger firms because of differential pricing.

3.3.3. Fixed product attributes. Finally, we explore the notion that large firms are large because they sell to richer households. If firms were born with fixed product attributes and/or brand perceptions, and some firms got lucky to appeal to the rich, while other producers cannot respond over time by altering their own product attributes or brand perceptions, this would mechanically lead to richer households sourcing from larger firms (as the rich account for a larger share of total sales).[18]

Here, we document that in the medium or long run this notion seems hard to reconcile with either the data or the existing literature on endogenous quality choice by firms. First, a body of empirical work has documented that firms endogenously choose their product attributes as a function of market demand in a variety of different empirical settings (*e.g.* Dingel, 2016; Bastos *et al.*, 2018).[19] Second, the scanner data suggest that producers of brands frequently alter the physical characteristics and/or presentation of their products over time. We find that during each half year close to 10% of producers of brands replace their products with changed product characteristics (*e.g.* packaging or product improvements) that have the identical pack sizes to the previous replaced varieties on offer by the same brand (Table A.4)—suggesting that producers are indeed capable of choosing their product attributes as a function of market conditions.[20] In

17. We condition on 79 store formats within the same county to capture potential differences in access across inner-city versus suburbs or due to car ownership. Conditioning on individual stores would give rise to the concern that households choose to shop at different retailers precisely due to the product mix on offer, rather than capturing differences in access.

18. This also relates to the original note in Melitz (2003) that the heterogeneity parameter can either be thought of as a marginal cost draw in a setting with horizontal differentiation, or as a quality draw in a setting with vertical differentiation.

19. Another literature in support of this is the marketing literature on firm strategies using advertising to affect brand perceptions over time (*e.g.* Keller *et al.*, 2011).

20. It could still be the case that our 18 repeated cross-sections (half years) depicted in Figure 1 are partly capturing the result of short-term taste shocks across products that differ between rich and poor households while hitting a fixed number of producers with fixed product attributes. To further investigate this possibility, we re-estimate the relationship in Figure 1 after replacing contemporary differences in firm sales by either sales three years before or three years in the future of the current period. If the distribution of firm sizes was subject to significant temporary swings over time, then we would expect the two counterfactual relationships to slope quite differently from our baseline estimate in Figure 1. Instead, Figure A.18 suggests that the estimated differences in producer sizes are practically identical.

support of these descriptive moments, we also provide more direct empirical evidence in Section 5 as part of our technology parameter estimation, documenting that an exogenous increase in the scale of production leads to brand-level quality upgrading over time (see Tables 2 and 3).

4. THEORETICAL FRAMEWORK

This section develops a tractable quantitative model that rationalizes the observed moments in the microdata. We introduce two basic features into an otherwise standard Melitz model of heterogeneous firms. On the demand side, we allow for non-homothetic preferences so that consumers across the income distribution can differ in both their price elasticities and in their product quality evaluations. On the producer side, firms with different productivities face the observed distribution of consumer preferences and optimally choose their product attributes and markups. Reflecting the asymmetry of market shares within sectors we document above, we also depart from the assumption of monopolistic competition, and allow for variable markups under oligopolistic competition. The exposition below abstracts from time subscripts. Supplementary Appendices 2–5 provide additional details on the model and its extensions.

4.1. *Model setup*

4.1.1. Consumption. The economy consists of two broad sectors: retail consumption (goods available in stores and supermarkets) and an outside sector. As in Handbury (2019), we consider a two-tier utility where the upper-tier depends on utility from retail shopping U_G and the consumption of an outside good z:

$$U = U(U_G(z), z). \tag{1}$$

For the sake of exposition, we do not explicitly specify the allocation of expenditures in retail versus non-retail items, but assume that the outside good is normal.[21] We denote by $H(z)$ the cumulative distribution of z across households in equilibrium, and normalize to one the population of consumers. Utility from retail consumption is then defined by:

$$U_G(z) = \prod_n \left[\sum_{i \in G_n} (q_{ni} \varphi_{ni}(z))^{\frac{\sigma_n(z)-1}{\sigma_n(z)}} \right]^{\alpha_n(z) \cdot \frac{\sigma_n(z)}{\sigma_n(z)-1}} \tag{2}$$

where n refers to a product module in the Nielsen data and i refers to a brand producer within the product module.[22] $\varphi_{ni}(z)$ is the perceived quality (product appeal) of brand i at income level z, and $\sigma_n(z)$ is the elasticity of substitution between brands in module n at income level z. Allowing demand parameters for retail consumption to be functions of outside good consumption z introduces non-homotheticities in a very flexible manner.[23] As we focus most of our attention

21. Supplementary Appendix 3 provides an envelope theorem result that holding z fixed yields a first-order approximation of the compensating variation from retail price shocks under any arbitrary (unspecified) upper-tier utility function. In the counterfactual analysis in Section 6, we report results both holding z fixed and allowing for endogenous changes in z after specifying (1).

22. We show in Supplementary Appendix 3.D that these preferences can be derived from the aggregation of discrete-choice preferences across many individual agents within group z.

23. For instance, demand systems with a choke price can generate price elasticities that depend on income (Arkolakis *et al.*, 2018), but offer significantly less flexibility in that relationship. Related work by Feenstra and Romalis (2014) features a single demand elasticity across incomes. Handbury (2019) derives conditions under which such preferences are rational and well-defined, and Supplementary Appendix 3.C provides an equivalent demand structure, based on results in Fally (2018), in which non-homotheticities do not rely on outside good consumption z.

on within-product module allocations, we model the choice over product modules with a Cobb–Douglas upper-tier, where $\alpha_n(z)$ refers to the fraction of expenditures spent on product module n at income level z (assuming $\sum_n \alpha_n(z)=1$ for all z).[24] Comparing two goods i and j within the same module n, relative expenditures by consumers of income level z are then given by:

$$\log\frac{x_{ni}(z)}{x_{nj}(z)} = (\sigma_n(z)-1)\left[\log\frac{\varphi_{ni}(z)}{\varphi_{nj}(z)} - \log\frac{p_{ni}}{p_{nj}}\right] \tag{3}$$

Motivated by the evidence discussed above, we let household quality evaluations $\log\varphi_{ni}(z)$ depend on an intrinsic quality term $\log\phi_{ni}$ associated with brand i and a multiplicative term $\gamma_n(z)$ depending on income level z:

$$\textit{Intrinsic Quality Assumption:} \qquad \log\varphi_{ni}(z)=\gamma_n(z)\log\phi_{ni}. \tag{4}$$

With the normalization $\int_{\Omega_z}\gamma_n(z)dz=1$ (where Ω_z is a set of z household types),[25] this intrinsic quality term also corresponds to the democratic average quality evaluation across households:

$$\log\phi_{ni} = \int_{\Omega_z}\log\varphi_{ni}(z)dz. \tag{5}$$

In the empirical estimation below, we estimate perceived quality $\varphi_{ni}(z)$ separately for each income group to verify whether relative quality evaluations are indeed preserved across income levels before imposing the above restriction. Finally, the retail price index is income-specific and given by $P_G(z)=\prod_n P_n(z)^{\alpha_n(z)}$, where the price index $P_n(z)$ for each product module n is defined as:

$$P_n(z)=\left[\sum_{i\in G_n} p_{ni}^{1-\sigma_n(z)}\varphi_{ni}(z)^{\sigma_n(z)-1}\right]^{\frac{1}{1-\sigma_n(z)}}. \tag{6}$$

4.1.2. Production. For each product group n, entrepreneurs draw their productivity a from a cumulative distribution $G_n(a)$ upon paying a sunk entry cost F_{nE}, as in Melitz (2003). For the remainder of this section, we index firms by a instead of i, since all relevant firm-level decisions are uniquely determined by firm productivity a. The timing of events is as follows. First, entrepreneurs pay the entry cost F_{nE} and discover their productivity a. Second, each entrepreneur decides at which level of quality to produce or exit. Third, production occurs and markets clear. Firms compete in prices, allowing for oligopolistic interactions as in Hottman *et al.* (2016).[26]

24. We abstract from within-brand product substitution by summing up sales across potentially multiple barcodes. Supplementary Appendix 4.A presents an extension of our model to multi-product firms following Hottman *et al.* (2016). In our setting, we show that as long as the ratio of between to within-brand elasticities of substitution does not significantly differ across income groups z, firm variation in within-brand product variety does not play a role for differences in firm sizes across z (Figure 1). In Section 5.1, we find minor differences in the cross-brand elasticities across income groups, and in Table A.5, we find no evidence of significant differences in the between-to-within ratios.

25. This normalization sets the simple mean of preference parameters $\gamma_n(z)$ equal to unity across a fixed set of household types z. As shown in Supplementary Appendix 3.E, this normalization to unity across household types z is without loss of generality.

26. We also consider a generalized version of Cournot quantity competition following Atkeson and Burstein (2008), whereby firms compete in quantity. This yields very similar results, both for the cross-section of implied markups as well as for the counterfactual analysis.

We normalize the cost of labour (wage w) to unity.[27] There are two cost components: a variable and a fixed cost (in terms of labour). We allow for the possibility that both the marginal and the fixed cost of production are a function of output quality. The latter captures potential overhead costs such as design, R&D, and marketing which do not directly depend on the quantities being produced but affect the quality of the product. In turn, variable costs depend on the level of quality of the production as well as the entrepreneur's productivity, as in Melitz (2003). Hence, the total cost associated with the production of a quantity q with quality ϕ and productivity a is:

$$c_n(\phi)q/a+f_n(\phi)+f_{0n}, \tag{7}$$

where $f_n(\phi)$ is the part of fixed costs that directly depend on output quality. For tractability, we adopt a simple log-linear parameterization for fixed costs as in Hallak and Sivadasan (2013) and the second model variant of Kugler and Verhoogen (2012):

$$f_n(\phi)=b_n\beta_n\phi^{\frac{1}{\beta_n}}. \tag{8}$$

For instance, fixed costs increase with quality ($\beta_n>0$) if higher quality entails higher expenditures on research and development or marketing, for which the cost is not directly dependent of the scale of production. On the contrary, it could be the case that fixed costs are decreasing in quality if such fixed investments are most suitable for mass-producing low-quality output at lower marginal costs.[28] Similarly, we assume that variable costs depend log-linearly on quality, with parameter $\xi_n\geq 0$ to capture the elasticity of the cost increase to the level of output quality:[29]

$$c_n(\phi)=\phi^{\xi_n}. \tag{9}$$

As long as ξ_n is smaller than the minimum quality evaluation $\gamma_n(z)$, firms choose positive levels of quality in equilibrium, as we further discuss below.

4.2. *Equilibrium and counterfactuals*

In equilibrium, consumers maximize their utility, expected profits upon entry equal the sunk entry cost, and firms choose their price, quality and quantity to maximize profits. There are two sources for variable markups across firms in our setting. They are determined by both their market power in oligopoly as well as the composition of the consumers that each firm sells to (*i.e.* the sales-weighted average price elasticity they face across consumer groups z). To see this, prices are given by:

$$p_n(a)=\frac{\phi(a)^{\xi_n}\mu_n(a)}{a}, \tag{10}$$

where μ_n is the markup over marginal cost:

$$\mu_n(a)\equiv\frac{p_n(a)}{c_n(a)}=1+\frac{\int_z x_n(z,a)dH(z)}{\int_z(\sigma_n(z)-1)(1-s_n(z,a))x_n(z,a)dH(z)} \tag{11}$$

27. Our focus is on the consumption side while shutting down potential implications of firm heterogeneity for nominal earnings inequality (*e.g.* Song *et al.*, 2018; Helpman *et al.*, 2017). In counterfactuals, GE labour market adjustments could have additional knock-on effects on relative prices across firms and, thus, differences in inflation between income groups, that our analysis ignores.

28. In Supplementary Appendix 4.B, we also present a model extension allowing for heterogeneous fixed costs that may be correlated with productivity, as in Hallak and Sivadasan (2013).

29. There is no need for a constant term as it would be isomorphic to a common productivity shifter after redefining $G_n(a)$.

$x_n(z,a)$ denotes sales of firm with productivity a to consumers of income level z. $s_n(z,a)$ are market shares in product module n. Due to oligopolistic competition, this markup is larger than $\frac{\tilde{\sigma}_n(a)}{\tilde{\sigma}_n(a)-1}$, the markup under monopolistic competition, where $\tilde{\sigma}_n(a)$ is the weighted average elasticity of substitution across consumers that the firm sells to:

$$\tilde{\sigma}_n(a)=\frac{\int_z \sigma_n(z)x_n(z,a)dH(z)}{\int_z x_n(z,a)dH(z)}.$$

The first-order condition in ϕ characterizes optimal quality $\phi_n(a)$ for firms with productivity a. Existence of a non-degenerate equilibrium choice for quality requires that we fit in either of two cases.[30] In the first case, higher quality entails higher fixed costs ($\beta_n>0$ and $b_n>0$) and the increase in marginal costs does not exceed consumer valuation for quality ($\xi_n<\gamma_n(z)$ for all z). In this case, large firms sort into producing higher quality. In the second case, marginal costs increase more strongly with higher quality output ($\xi_n>\gamma_n(z)$ for all z), and fixed costs are decreasing ($\beta_n<0$ and $b_n<0$). In this case, relatively small "boutique" producers sort into producing higher quality products while mass producers produce low-cost, low-quality products.

In both cases, optimal quality satisfies:

$$\phi_n(a)=\left(\frac{\tilde{\gamma}_n(a)-\xi_n}{b_n\mu_n(a)}.X_n(a)\right)^{\beta_n}, \tag{12}$$

where $X_n(a)=\int_z x(a,z)dH(z)$ denotes total sales of firm a in product module n and $\tilde{\gamma}_n(a)$ is the weighted average quality valuation $\gamma_n(z)$ for firm with productivity a, weighted by sales and price elasticities across its customer base:

$$\tilde{\gamma}_n(a)=\frac{\int_z \gamma_n(z)(\sigma_n(z)-1)(1-s_n(z,a))x_n(z,a)dH(z)}{\int_z(\sigma_n(z)-1)(1-s_n(z,a))x_n(z,a)dH(z)}. \tag{13}$$

Optimal quality is determined by several forces that are apparent in equation (12). First, when $\beta_n>0$, larger sales induce higher optimal quality, as reflected in the term $X_n(a)^{\beta_n}$. This is the scale effect due to the higher fixed costs of producing at higher quality. If we compare two firms with the same customer base, the larger one would more profitably invest in upgrading quality if $\beta_n>0$. Second, optimal quality depends on how much the firm-specific customer base value quality, captured by $\tilde{\gamma}_n(a)$. Firms that tend to sell to consumers with high $\gamma_n(z)$ also tend to have higher returns to quality upgrading. Third, optimal quality depends on technology and the cost structure. A higher elasticity of marginal costs to quality, ξ_n, induces lower optimal quality. If instead $\beta_n<0$, higher quality is associated with smaller firm scale $X_n(a)$, a higher average quality valuation $\tilde{\gamma}_n(a)$ and a lower elasticity of marginal costs to quality ξ_n.

Finally, when firms chose prices and quality to maximize profits, those profits are given by:

$$\pi_n(a)=\left(1-\frac{1}{\mu_n(a)}\right)\left[\int_z\big(1-\beta_n(\gamma_n(z)-\xi_n)(\sigma_n(z)-1)(1-s_n(z,a))\big)x_n(a,z)dH(z)\right]-f_{0n} \tag{14}$$

where $\beta_n(\gamma_n(z)-\xi_n)(\sigma_n(z)-1)(1-s_n(z,a))$ is the share of operating profits that are invested in the fixed costs of quality upgrading.[31] In equilibrium, average profits (assuming equality between

30. If $\beta_n>0$, $b_n>0$, and $\xi_n>\gamma_n(z)$, optimal quality is zero. If $\beta_n<0$, $b_n<0$, and $\xi_n<\gamma_n(z)$, optimal quality is infinite.

31. This term must be smaller than unity. Second-order conditions require $\beta_n(\gamma_n(z)-\xi_n)(\sigma_n(z)-1)<1$ to ensure a well-defined equilibrium for all firms.

averages and expectations) across all entrants equal sunk entry costs, and surviving firms are those with positive profits.

An equilibrium is defined as a set of sales, quality choices and markups that satisfy equations (3), (10), (12), and (11) for each firm, such that price indices are given by (6) and such that profits are given by (14). Additional details are provided in Supplementary Appendix 2.[32]

4.2.1. Firm heterogeneity across consumption baskets. To rationalize the observed stylized facts through the lens of the model, we examine the weighted average of log firm size $X_n(a)$ for each income group z, which corresponds to what we plot on the y-axis of Figure 1:

$$\log\widetilde{X}_n(z) = \frac{\int_a x_n(z,a)\log X_n(a)dG_n(a)}{\int_a x_n(z,a)dG_n(a)}.$$

How $\widetilde{X}_n(z)$ varies with income (*i.e.* the slope of the estimated relationship in Figure 1) reflects how $x_{ni}(z,a)$ varies across firms i and consumer income z. For the sake of exposition, let us assume for now that quality valuation $\gamma_n(z)$ and price elasticities $\sigma_n(z)$ are continuous and differentiable w.r.t income z. We can then express the derivative $\frac{\partial \log\widetilde{X}_n(z)}{\partial z}$ as a function of two covariance terms (where Cov_z denotes a covariance weighted by sales to consumers z):

$$\begin{aligned} \frac{\partial \log\widetilde{X}_n(z)}{\partial z} = & \frac{\partial \gamma_n(z)}{\partial z}(\sigma_n(z)-1)\, Cov_z(\log X_n(a), \log\phi_n(a)) \\ & -\frac{\partial \sigma_n(z)}{\partial z}\, Cov_z\Big(\log X_n(a), \log(p_n(a)/\phi_n(a)^{\gamma_n(z)})\Big). \end{aligned} \tag{15}$$

From this expression, we see that the difference in weighted-average firm size in consumption baskets across the income distribution is driven by how preference parameters depend on income ($\frac{\partial\gamma_n}{\partial z}$ and $\frac{\partial\sigma_n}{\partial z}$), and by how firm size correlates with quality and quality-adjusted prices. The first line in equation (15) reflects a quality channel. It is positive if firm size increases with quality and if richer households care relatively more about intrinsic product quality ($\frac{\partial\gamma_n}{\partial z} > 0$). The second term captures a price effect, which would work in the same direction as the quality channel if, and only if, richer households were more price elastic compared to poorer households, as the final covariance term between firm size and quality-adjusted prices is negative (lower quality-adjust prices lead to larger sales when $\sigma_n(z) > 1$). If, instead, higher income consumers were less price elastic, but attached greater value to product quality, the two channels in (15) would be opposing one another in generating the observed heterogeneity in firm sizes across income groups in Figure 1.

The decomposition in equation (15) relies primarily on our demand-side structure and does not yet impose assumptions on the production side. In turn, the supply-side structure can shed light on the potential sources of the covariance terms. Prices are given by equation (10) while equilibrium product quality satisfies equation (12). In particular, the correlation between firm size and quality appearing in the first term can be expressed as:

$$Cov_z(\log X_n(a), \log\phi_n(a)) = \beta_n Var_z(\log X_n(a)) + \beta_n Cov_z(\log X_n(a), \log((\tilde{\gamma}_n(a)-\xi_n)/\mu_n(a))). \tag{16}$$

32. Supplementary Appendix 2 also examines second-order conditions, discusses conditions for uniqueness, and describes a special case with closed-form solutions.

As part of the estimation that follows, we can quantify each of these terms and decompose the observed firm heterogeneity across the consumption baskets in Figure 1 into the underlying channels.

4.2.2. Counterfactual analysis. Our framework naturally lends itself to quantifying GE counterfactuals. In Supplementary Appendix 5.A, we write the equilibrium equations in terms of counterfactual changes that govern adjustments in firm sales, quality, variable markups, entry, exit, and price indices. Solving for counterfactual equilibria requires data on initial sales $x_{n0}(z,a)$ for each firm across different consumer groups, in addition to estimates of demand and supply parameters: $\sigma_n(z)$, $\gamma_n(z)$, β_n, and ξ_n. With these moments in hand, we can solve for changes in quality $\frac{\phi_{n1}(a)}{\phi_{n0}(a)}$, sales $\frac{x_{n1}(z,a)}{x_{n0}(z,a)}$, markups $\frac{\mu_{n1}(a)}{\mu_{n0}(a)}$, the mass of firms $\frac{N_{n1}}{N_{n0}}$, firm survival $\delta_{nD}(a)$, and consumer price indices $\frac{P_{n1}(z)}{P_{n0}(z)}$. In particular, equilibrium changes in quality can be derived by taking ratios of equation (12), changes in sales are derived from equations (3) and (10), changes in markups from equation (11), changes in profits from equation (14), and changes in cost of living from equation (6). As described in Supplementary Appendix 5.A, we do not require estimates of firm productivity a or initial firm quality $\phi(a)$ to conduct our counterfactual exercise.[33] Supplementary Appendix 5.B also derives a 6-fold decomposition of the counterfactual price index effects that we use and further discuss as part of the counterfactual analysis in Section 6.

In our baseline counterfactual analysis, we hold initial household types z fixed, without specifying the upper-tier utility function in (1).[34] As shown in Supplementary Appendix 3, this provides a first-order approximation of the compensating variation due to retail price changes for any arbitrary upper-tier utility. To investigate potential second-order effects on price indices (through changes in demand parameters $\alpha_n(z)$, $\sigma_n(z)$, and $\gamma_n(z)$), we then also examine counterfactuals allowing for endogenous changes in z after specifying the upper-tier in (1).

5. ESTIMATION

This section presents parameter estimation and model calibration. To bring the model to the data, we define five consumer income groups z in terms of quintiles of the U.S. income distribution, and introduce time subscripts to reflect 18 half-year cross-sections of data. We begin by estimating the preference parameters, σ_{nz} and γ_{nz}, that combined with detailed data on sales and unit values, allow us to quantify the distribution of product quality, quality-adjusted prices and markups across producers of brands and household consumption baskets. With these estimates in hand, we then proceed to estimate the technology parameters, β_n and ξ_n. As well as being of interest in their own right, these estimates allow us to quantify the forces underlying Figure 1 at the end of this section, and to explore policy counterfactuals in the final section.

5.1. *Price elasticities*

We begin by estimating the demand elasticity $(1-\sigma_{nz})$ that we allow to vary across household income groups and product groups. From equation (3), we get the following estimation equation:

$$\triangle \log\left(s_{nzict}\right) = (1-\sigma_{nz})\triangle \log\left(p_{nict}\right) + \eta_{nzct} + \epsilon_{nzict}, \tag{17}$$

33. This approach follows Dekle *et al.* (2007) among others.

34. This is similar to *e.g.* Handbury (2019), Atkin *et al.* (2018), and Redding and Weinstein (2020) and follows earlier work by McFadden and Train (2000).

TABLE 1
Price elasticities

Panel A: Pooled estimates Dependent variable: change in log budget shares	OLS	National IV	State IV	Both IVs	Both IVs
(1-σ) All households	0.257***	−1.184***	−1.090***	−1.181***	
	(0.0288)	(0.0356)	(0.0415)	(0.0316)	
(1-σ) Poorest quintile (relative to richest)					−0.375***
					(0.109)
(1-σ) 2nd poorest quintile (relative to richest)					−0.391***
					(0.0957)
(1-σ) median quintile (relative to richest)					−0.163***
					(0.0558)
(1-σ) 2nd richest quintile (relative to richest)					−0.271***
					(0.0862)
Quintile-by-module-by-county-by-period FX	✓	✓	✓	✓	✓
Brand-by-county-by-period FX	×	×	×	×	✓
Observations	9,989,508	9,989,508	9,283,699	9,283,699	9,283,699
First stage F-stat		718.6	314.7	420.0	123.0

Panel B: By product department Dependent variable: change in log budget shares	Beverages Both IVs	Dairy Both IVs	Dry grocery Both IVs	Frozen foods Both IVs	General merchandize Both IVs	Health and beauty Both IVs	Non-food grocery Both IVs	Packaged meat Both IVs
(1-σ) All households	−1.091***	−0.716***	−1.324***	−1.336***	−2.353***	−0.504***	−1.100***	−1.318***
	(0.149)	(0.0559)	(0.0405)	(0.0672)	(0.222)	(0.0878)	(0.0911)	(0.151)
Quintile-by-module-by-county-by-period FX	✓	✓	✓	✓	✓	✓	✓	✓
Observations	755,648	775,238	4,570,372	945,956	205,830	778,667	982,261	269,726
First stage F-stat	542.2	253.0	407.8	126.7	169.2	217.0	731.7	56.63

Panel C: By department and household group Dependent variable: change in log budget shares	Beverages Both IVs	Dairy Both IVs	Dry grocery Both IVs	Frozen foods Both IVs	General merchandize Both IVs	Health and beauty Both IVs	Non-food grocery Both IVs	Packaged meat Both IVs
(1-σ) below median quintiles	−1.272***	−0.809***	−1.481***	−1.341***	−2.436***	−0.506*	−1.383***	−1.329***
	(0.252)	(0.142)	(0.105)	(0.148)	(0.368)	(0.272)	(0.239)	(0.261)
(1-σ) median and above quintiles	−1.041***	−0.689***	−1.288***	−1.336***	−2.339***	−0.501***	−1.048***	−1.316***
	(0.147)	(0.0569)	(0.0462)	(0.0721)	(0.249)	(0.107)	(0.0757)	(0.155)
Quintile-by-module-by-county-by-period FX	✓	✓	✓	✓	✓	✓	✓	✓
Observations (when pooled)	755,648	775,238	4,570,372	945,956	205,830	778,667	982,261	269,726

Notes: See Section 5 for discussion. Columns in Panel C report point estimates from two separate regressions. Standard errors are in parentheses below point estimates and clustered at the level of counties. ***, **, * indicate 1, 5, and 10% confidence levels.

where, as before, z, n and i denote household groups, product modules and brands. c and t indicate U.S. counties and 18 half years (17 changes), and s_{nzict} are budget shares within product module n. η_{nzct} are household group-by-product module-by-county-by-half-year fixed effects that capture the local CES price index. Consistent with our specification of preferences at the level of household groups, we estimate (17) after aggregating consumption shares in the home scanner data for the period 2006–14 to the level of household quintile-by-county-by-module-by-half-year bins.[35] To address concerns about autocorrelation in the error term ϵ_{nzict}, we cluster standard errors at the county level.[36]

To address the standard simultaneity concern that taste shocks in the error term are correlated with observed price changes, we follow the empirical literature in industrial organization (e.g. Hausman (1999), Nevo (2000), and Hausman and Leibtag (2007)) and make the identifying assumption that consumer taste shocks are idiosyncratic across counties whereas supply-side cost shocks are correlated across space. For the supply-side variation needed to identify σ_{nz}, we exploit the fact that store chains frequently price nationally or regionally without taking into consideration changes in local demand conditions (DellaVigna and Gentzkow, 2019). In particular, we instrument for local consumer price changes across brands $\triangle \log(p_{nict})$ with either national or state-level leave-out mean price changes: $\frac{1}{N-1}\sum_{j\neq c}\triangle \log\left(p_{nijt}\right)$. These two instruments identify potentially different local average treatment effects. The national leave-out means IV estimates σ_{nz} off retail chains that price their products nationally, whereas the state-level leave-out means may extend the complier group to regional and local retailers.

A potentially remaining concern that this IV strategy would not be able to address are demand shocks at the national or state-level that are correlated with observed product price changes. Advertisement campaigns would be a natural candidate for this concern. For this to lead to a downward bias in the σ_{nz} estimates, it would have to be the case that the advertisement campaign first affects demand, but then also leads to higher prices. We would argue that this is not likely to be the case for most national or state-level advertisement campaigns. For example, an "informative" advertisement campaign containing price information would not lead to a bias in our estimation of σ_{nz}, as the variation is driven by consumers reacting to a change in prices (promotions etc.). A second type of "persuasive" campaign could be aimed at improving the brand's perception instead, which would be more problematic for the exogeneity of the IV. For identification, we require that it is not the case that firms on average launch persuasive advertisement campaigns and simultaneously increase their prices. Given the longer-term objective of most image-oriented advertisement campaigns (*e.g.* Keller *et al.*, 2011), and the fact that we use half-yearly variation in prices and consumption decisions in our estimations, we believe this to be a plausible baseline assumption.

To document sensitivity, we also report counterfactual results in Section 6 across alternative parameter assumptions. Finally, the key empirical moment in our welfare quantification does not rely on the levels of σ_{nz}, but on the observed heterogeneity across different income groups. And while it is possible that some of the discussed endogeneity concerns may affect rich and poor

35. We aggregate household purchases to income groups using sampling weights ("projection factor" in Nielsen) to compute $\triangle \log(s_{nzict})$, and limit the sample to income group-by-county-by-half-year cells with at least 25 households. To compute brand-level log price changes we first compute projection-factor-weighted price means for each barcode-by-county-by-half-year cell, and then compute $\triangle \log(p_{nict})$ as a brand-level Tornqvist price index across all barcodes belonging to the same brand. As reported in Table A.6, point estimates are not sensitive to either the decision to take mean prices (rather than medians) or the decision to take a Tornqvist price index (rather than Laspeyres or a simple average).

36. Clustering at this level yields slightly more conservative standard errors than sensible alternatives (clustering at the level of brands, product modules, county-by-income groups, county-by-half-years, or county-by-product modules).

households differently, such concerns would require somewhat more elaborate stories compared to the traditional simultaneity bias in demand estimation.[37]

5.1.1. Estimation results. Table 1 panel A shows the pooled estimation results across all household and product groups. In support of the IV strategy, we find that the point estimates change from slightly positive in the OLS specification to negative and statistically significant in both IV estimations as well as the joint IV column. The estimates from the two different instruments are very similar and suggest a pooled elasticity of substitution of about 2.2. Table 1 panel B presents IV estimates separately across product departments that range between 1.5 and 3.5. These estimates fall at the centre of a large existing literature in IO and Marketing using brand-level consumption data to estimate the sales-to-price elasticity of demand.[38] They are, however, somewhat lower than empirical work based on moment conditions of the double-differenced residuals in demand and supply using the GMM estimation approach pioneered by Feenstra (1994) (*e.g.* Broda and Weinstein, 2010; Hottman *et al.*, 2016).[39] As a robustness exercise, we report counterfactuals in the final section for both our baseline estimates and assuming larger values of the price elasticity.

In the final column of Table 1 panel A, we take the pooled sample, but interact the log price changes with household income group identifiers to estimate to what extent there are statistically significant differences between household quintiles. The most convincing way to estimate such household differences in σ_{nz} is to additionally include brand-by-period-by-county fixed effects, so that we identify differences in the elasticity of substitution by comparing how different households react to the identical price change. We choose the richest income group as the reference category absorbed by the additional fixed effects. We find that poorer households have significantly higher price elasticities compared to richer households. In terms of magnitude, however, these differences are relatively minor. We estimate that the price elasticity for the poorest two quintiles is about 0.4 larger than that for the richest quintile.

Table 1 panel C reports results for each of the product departments across two income groups: the bottom two quintiles and the top 3 quintiles. These sixteen σ_{nz} estimates reported in Table 1 panel C are the point estimates that we use as our baseline parameter values in the analysis that follows. This is motivated by the income group heterogeneity reported in the final column of Table 1 panel A, and due to the fact that statistical power starts to become an issue when estimating these parameters separately across individual product departments. The trade-off that we face here is one between relatively precisely estimated point estimates relative to allowing for richer patterns of heterogeneity. For completeness, we also report the results when estimating forty σ_{nz} parameters (5 across each of the 8 product departments) (Table A.7). A larger number of parameters start having large standard errors and lack statistical significance compared to our preferred set of estimates in Table 1 panel C. An alternative approach is to estimate the heterogeneity parametrically by interacting changes in log prices with either the average of log household total expenditure per capita in a given zct bin or the average percentile of total per capita expenditure in that bin (Table A.8). We use these alternative estimates as part of the sensitivity analysis in the counterfactual analysis below.

37. Also note that in the absence of additional linked household data on times spent shopping, it would be impossible to disentangle the potential for differences in the opportunity cost of search/shopping time across rich and poor households (*e.g.* Aguiar and Hurst, 2007).

38. See *e.g.* a meta-analysis by Bijmolt *et al.* (2005). Reviewing close to 2000 estimates, they find a median elasticity of −2.2, with one half of the estimates falling in the range between −1 and −3.

39. See also Soderbery (2015) and more recent work by Ray (2019) addressing potential upward bias due to many weak instruments in this setting.

5.2. *Brand quality, quality-adjusted prices, and markups*

Armed with estimates of σ_{nz}, we can use the scanner data following equations (3) and (5) to estimate product quality evaluations $\log\varphi_{nzi}$, $\log\phi_{ni} = \frac{1}{N_z}\sum_z \log\varphi_{nzi}$ and quality-adjusted prices, $\log\left(\frac{p_{ni}}{\phi_{ni}}\right)$, across producers of brands and household consumption baskets. To do this, we use an additional empirical moment from the data, product unit values, in combination with observed product sales and the estimated σ_{nz}.

We plot the distribution of mean deviations in log product unit values within product module-by-half-year cells across the income distribution (aggregated as expenditure-weighted averages for each household) (Figure A.19).[40] The richest quintile of US households source their consumption from firms that have on average 12% higher unit values within product modules compared to the poorest quintile.

The left panel of Figure 2 proceeds to present the distribution of the estimated weighted average product quality deviations across household consumption baskets. We find that the documented differences in terms of firm sizes translate into significant differences in the weighted average product quality as well as quality-adjusted prices embodied in consumption baskets across the income distribution. The richest 20% of US households source their consumption from on average 22% higher quality producers compared to the poorest 20% of households. Using the estimates of income-group-specific product quality shifters, we confirm the motivating evidence above (Figures A.6 and A.21): rich and poor households agree that product quality increases across the firm size distribution, but this relationship is steeper among richer households. Moving from differences in product quality to quality-adjusted prices, the right panel of Figure 2 documents that the richest quintile source their consumption at on average 10% lower quality-adjusted prices.

The parameter estimates for σ_{nz} in combination with data on firm sales by income group also allow us to compute the distribution of the effective (weighted average) elasticities of substitution faced by individual producers, $\left(\tilde{\sigma}_{ni} = \frac{\sum_z \sigma_{nz} x_{nzi}}{\sum_z x_{nzi}}\right)$. Following equation (11), this together with firm market shares under oligopolistic competition determines the distribution of markups across firms. The left panel of Figure 3 presents the estimation results for variation in μ_{ni} as a function of firm size deviations within product modules. Larger firms charge significantly higher markups due to both their market power within product modules and the fact that they face lower price elasticities due to selling a higher share of their output to higher-income households (who, in turn, have lower parameter values for σ_{nz}).

With estimates of φ_{nzi} in hand, we proceed to estimate the final set of preference parameters, γ_{nz}, that govern the valuation of product quality characteristics across the household income distribution. Following expression (4), the estimation equation is:

$$\log\left(\varphi_{nzit}\right) = \gamma_{nz}\log\left(\phi_{nit}\right) + \eta_{nzt} + \epsilon_{nzit}, \tag{18}$$

where η_{nzt} are income group-by-product module-by-half-year fixed effects. To address the concern of correlated measurement errors when moving from model to data, that appear both on the left hand side (the income group specific product quality evaluations) and the right hand side (the democratic average product quality evaluation), we instrument for $\log(\phi_{nit})$ with two half-year lagged values of product quality. To address autocorrelation in the error term ϵ_{nzit}, we cluster standard errors at the level of product modules.

Table A.9 presents the estimation results across bins of household groups and product departments. In accordance with the documented stylized facts, richer household groups value

40. We compute brand-level unit values as sales-weighted means across barcode transactions at the level of brands-by-half-year cells. Figure A.20 depicts the relationship between firm sizes and unit values.

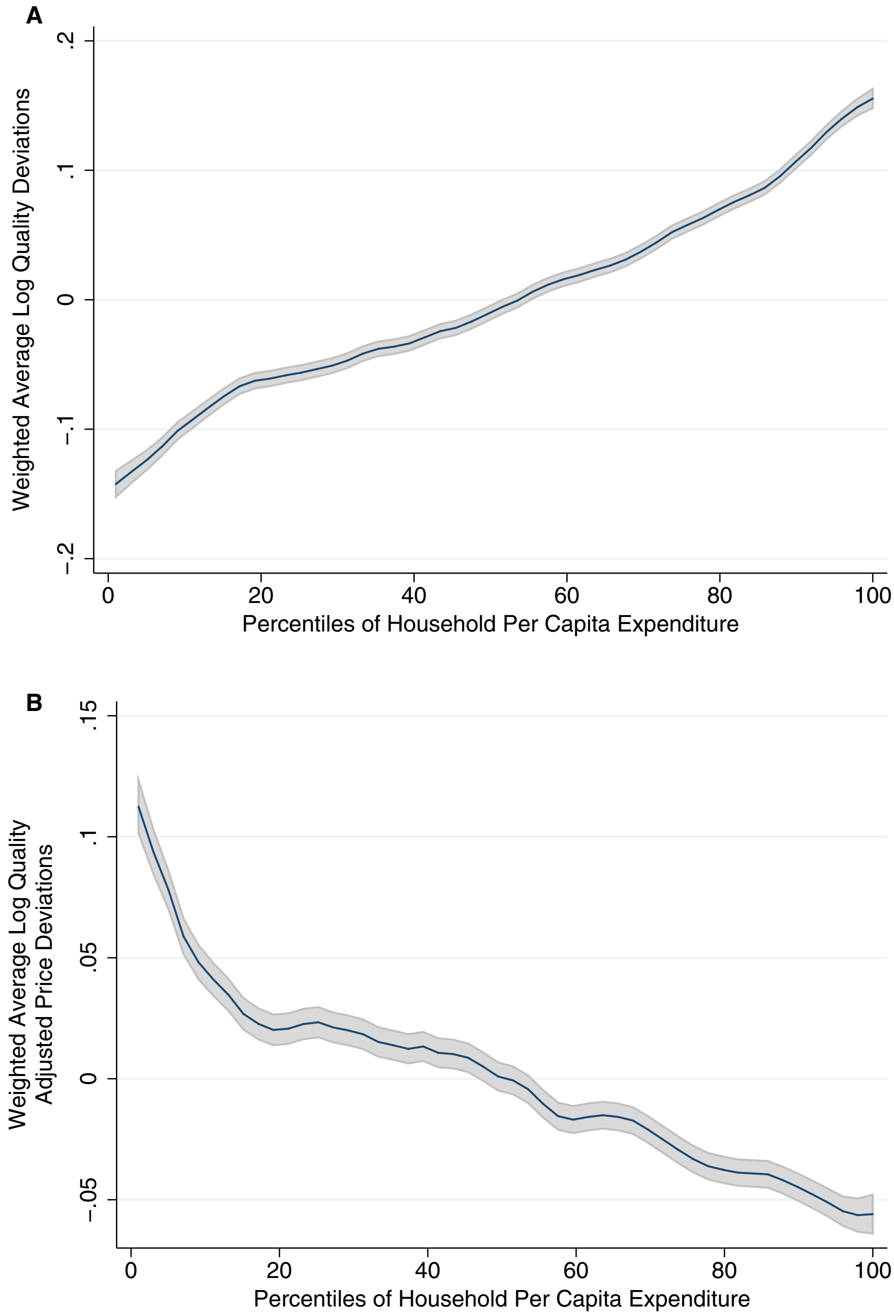


FIGURE 2

Distribution of weighted average product quality and quality-adjusted prices across consumption baskets

Notes: The figure depicts deviations in weighted average log brand quality embodied in the consumption baskets of on average 59,000 US households during 18 half-year periods between 2006 and 2014. The y-axis in Panel A (Panel B) displays weighted average deviations in log brand quality (quality-adjusted prices) within more than 1000 product modules where the weights are household expenditure shares across producers of brands. The x-axis in both graphs displays national percentiles of per capita total household retail expenditure per half-year period (see Section 2). The fitted relationships correspond to local polynomial regressions. Standard errors in both graphs are clustered at the county level, and the displayed confidence intervals are at the 95% level. See Section 5 for discussion.

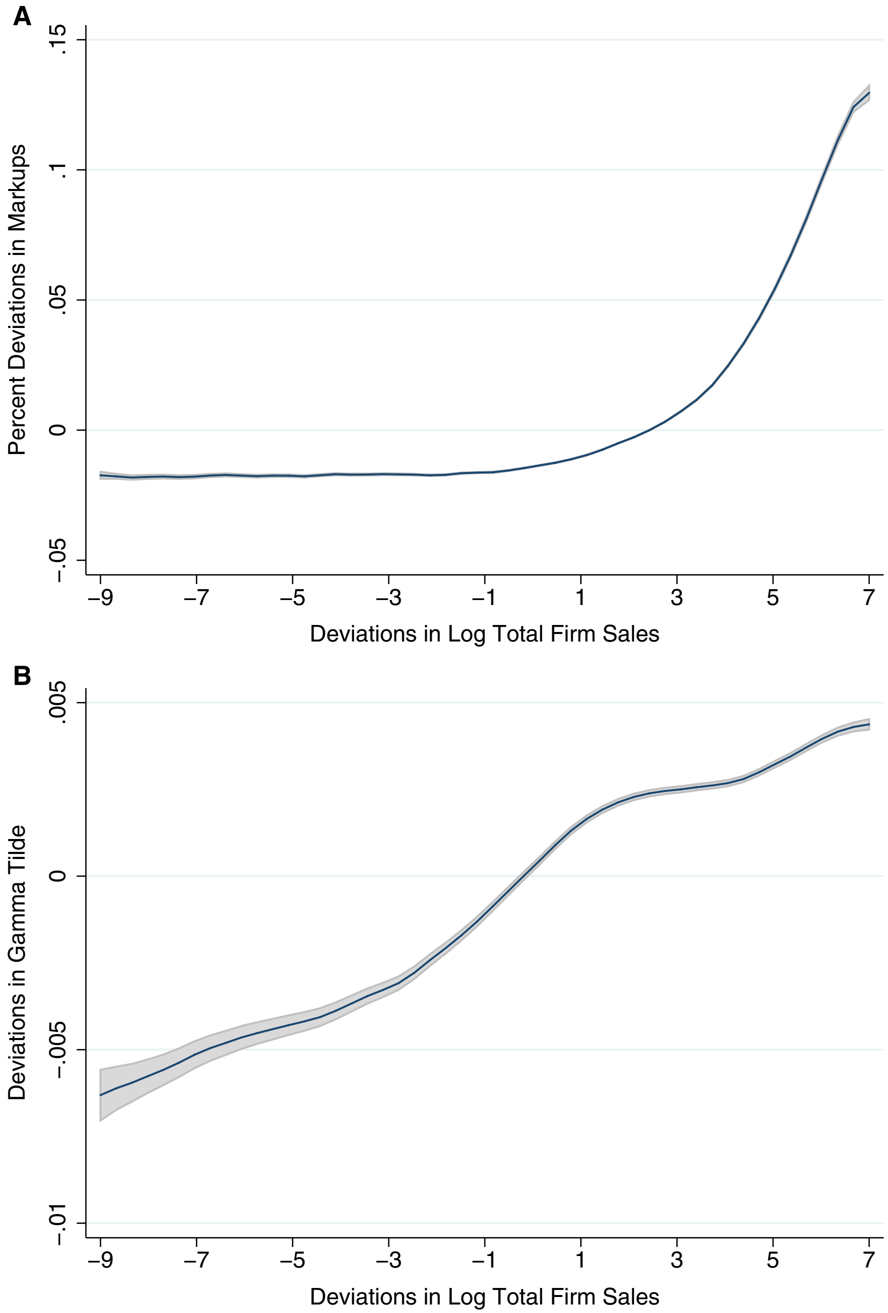


FIGURE 3

Producers charge variable markups and face different tastes for quality

Notes: The figure depicts deviations in the markup (μ_{ni}) and taste-for-quality parameters ($\tilde{\gamma}_{ni}$) across the firm size distribution for 18 half-yearly cross-sections between 2006 and 2014. The y-axis displays deviations in μ_{ni} and $\tilde{\gamma}_{ni}$ relative to product module-by-half-year means. The x-axis displays deviations of log firm sales at the same level (between 1st and 99th percentiles for legibility). The fitted relationships correspond to local polynomial regressions. Standard errors in both graphs are clustered at the level of product modules, and the displayed confidence intervals are at the 95% level. See Section 5 for discussion.

increases in product quality relatively more across each of the product departments. However, there we also find differences in the extent of this heterogeneity across different product departments. For example, beverages, dairy products and packaged meat are among the departments with the highest difference in the taste for quality, whereas general merchandise and health and beauty care have the lowest differences in household taste for quality across income groups.

As we do above for the firm-level parameter $\tilde{\sigma}_{ni}$, we can use these estimates in combination with the sales data to compute the weighted average product quality evaluations faced by each brand producer, following expression (13). The right panel of Figure 3 reports these results across the firm size distribution within product groups. We find that larger producers of brands face a composition of market demand with significantly higher marginal valuations for product quality.

5.3. *Technology parameters*

In this subsection, we propose two approaches to estimate the technology parameters.

5.3.1. Estimation in the cross-section.

Armed with estimates of μ_{ni} and $\tilde{\gamma}_{ni}$, we proceed to estimate the technology parameters β_n and ξ_n: the first determines the presence and size of economies of scale in the production of product quality. The second determines the extent to which marginal costs increase with higher product quality. An intuitive way to estimate β_n is through the relationship between unit values and market shares within product modules. If we imposed the assumptions of monopolistic competition and homogeneous consumer preferences (representative agent), we would get the following estimation equation from (3) and (12) above:

$$\log(p_{nit}) = \left(\beta_n - \frac{1}{\sigma_n - 1}\right)\log(X_{nit}) + \eta_{nt} + \epsilon_{nit} \tag{19}$$

where η_{nt} are product module-by-half-year fixed effects. Intuitively, if brands were of the same quality then the relationship between unit values (that would be identical to prices in this case) and market shares would be governed by the slope of the demand curve $-\frac{1}{\sigma_n - 1}$. The extent to which firms of larger scale sort into producing higher product quality would then be captured by the production function parameter β_n. To see this more clearly, we can re-write (19) with product quality on the left hand side: $\log(\phi_{nit}) = \beta_n \log(X_{nit}) + \eta_{nt} + \epsilon_{nit}$, where following (3) and (5) $\log(\phi_{nit}) = \log(p_{nit}) + \frac{1}{\sigma_n - 1}\log(X_{nit})$. This same logic and estimation equation have been used in the existing literature on quality choice across heterogeneous firms under the representative agent assumption (*e.g.* Kugler and Verhoogen, 2012).

When allowing for variable markups in oligopolistic competition and heterogeneity in tastes for quality and price elasticities across consumers—giving rise to firm-specific demand compositions for $\tilde{\gamma}_{ni}$ and $\tilde{\sigma}_{ni}$—this estimation equation requires two additional correction terms. From (5) and (12) we get:

$$\log(p_{nit}) = \left(\beta_n - \frac{1}{\overline{\sigma}_n - 1}\right)\log(X_{nit}) - \frac{1}{N_z}\sum_z \frac{1}{\sigma_{nz} - 1}\log\left(\frac{X_{nzit}}{X_{nit}}\right) \tag{20}$$
$$+\beta_n \log((\tilde{\gamma}_{nit} - \xi_n)/\mu_{nit}) + \eta_{nt} + \epsilon_{nit},$$

TABLE 2
Product quality and firm scale: reduced-form evidence

	All product groups							
	Cross-section				Panel data			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log unit value		Log quality		Δ log unit value		Δ log quality	
Dependent variables:	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Log national firm sales	0.0280***	0.0253***	1.128***	1.142***				
	(0.00339)	(0.00390)	(0.0312)	(0.0309)				
Δ log national firm sales					0.0365***	0.0705***	1.131***	0.569***
					(0.00320)	(0.0138)	(0.0415)	(0.0589)
Product module-by-period FX	✓	✓	✓	✓	×	×	×	×
State-by-product module-by-period FX	×	×	×	×	✓	✓	✓	✓
Observations	1,330,976	1,330,976	1,330,976	1,330,976	1,789,078	1,789,078	1,789,078	1,789,078
Number of product module clusters	1,031	1,031	1,031	1,031	1,004	1,004	1,004	1,004
First stage F-stat		322,552		322,552	251.1		251.1	

Notes: See Section 5 for discussion. Standard errors are in parentheses below point estimates and clustered at the level of product modules. ***, **, * indicate 1, 5, and 10% confidence levels.

TABLE 3
Technology parameter estimates

Dependent variable:	All product groups			
Log product quality or	Cross-section		Panel data	
changes in log quality	OLS	IV	OLS	IV
Log firm scale or changes in log firm scale (β)	1.1132***	1.1352***	1.1637***	0.4886***
	(0.0309)	(0.0307)	(0.0466)	(0.0652)
ξ Parameter	0.82	0.82	0.35	0.35
Observations	1,330,976	1,330,976	1,422,244	1,422,244
Number of clusters	1,031	1,031	994	994
First stage F-stat	311, 103.21		247.73	

Dependent variable:	Grocery				Non-grocery			
Log product quality or	Cross-section		Panel data		Cross-section		Panel data	
changes in log quality	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Log firm scale or changes in log firm scale (β)	0.9318***	0.9501***	0.9408***	0.2914***	1.5328***	1.5767***	1.5296***	0.9531***
	(0.012)	(0.0127)	(0.0114)	(0.0748)	(0.085)	(0.0839)	(0.1008)	(0.1349)
ξ Parameter	0.81	0.81	0.06	0.06	0.87	0.87	0.65	0.65
Observations	1,002,542	1,002,542	1,031,295	1,031,295	328,434	328,434	390,949	390,949
Number of clusters	719	719	696	696	312	312	298	298
First stage F-stat		273,413.55		180.05		65,882.01		103.64

Notes: See Section 5 for discussion. Standard errors are in parentheses below point estimates and clustered at the level of product modules. ***, **, * indicate 1, 5, and 10% confidence levels.

where N_z is the number of consumer groups (5 in our application), $\frac{1}{\bar{\sigma}_n - 1} = \frac{1}{N_z}\sum_z \frac{1}{\sigma_{nz}-1}$. The first additional term on the right generalizes the downward-sloping demand relationship $\left(-\frac{1}{\bar{\sigma}_n - 1}\log(X_{nit})\right)$ in equation (19), to allow for the fact that different producers face different market demand elasticities due to differences in the composition of their customers. The second additional term captures the fact that different producers may sort into higher or lower product quality due to variable markups and differences in the composition of their customer base (valuing quality more or less given prices).

As above, using (3) and (5), we can re-write equation (20) for estimation as: $\log(\phi_{nit})= \beta_n \log(X_{nit}(\tilde{\gamma}_{nit}-\xi_n)/\mu_{nit})+\eta_{nt}+\epsilon_{nit}$. Conditional on η_{nt} fixed effects, this allows us to jointly estimate the technology parameters β_n and ξ_n for each product department by estimating β_n using IV regressions across iterations of ξ_n, and selecting the best-fitting parameter combination. We use iterations of ξ_n in steps of 0.01 in the range between 0 and twice the maximum estimated γ_{nz}, and do not impose an ex ante assumption about the existence of economies of scale in quality production ($\beta_n > 0$).

Two identification concerns in (20) are correlated measurement errors on the left and right hand sides, and temporary consumer taste shocks: deviations around ϕ_{ni} over time that would mechanically lead to a biased estimate $\beta_n = \frac{1}{\bar{\sigma}_n - 1}$ if unit values and firm quality (but not sales) remain unchanged in response to the temporary taste shock. To address both of these concerns, we instrument for markup and composition-adjusted firm scale $\log(X_{nit}(\tilde{\gamma}_{nit}-\xi_n)/\mu_{nit})$ with two half-year lags. To address concerns about autocorrelation in the error term, we cluster the standard errors at the level of product modules as before. A deeper concern with the cross-sectional estimation is that omitted factors, such as firm-specific quality upgrading costs discussed in Section 4 and Supplementary Appendix 4.B, that affect both firm scale and product quality choices are difficult to rule out.

Finally, cross-sectional variation would be insufficient to distinguish between alternative microfoundations to rationalize the observed sorting of larger firms into higher output quality. In our model we follow earlier work by *e.g.* Sutton (1998) allowing fixed costs to increase with output quality. Alternatively, *e.g.* Baldwin and Harrigan (2011) and Feenstra and Romalis (2014) model quality choices as a direct function of a firm's productivity draw. While both model variants give rise to a cross-sectional relationship between firm scale and product quality, only the former would be consistent with changes in output quality as a function of exogenous changes in firm scale in firm-level panel data.

5.3.2. Panel estimation. Estimation equation (20) extends the existing literature on quality choice across firms to a setting that also allows for heterogeneity on the consumption side. But it also follows the existing literature in that it is based on cross-sectional variation across firms. An alternative approach is to use within-brand variation over time. The natural panel data approach would be to write (20) in log changes instead of log levels on both the left and right hand sides. However, the estimation of β_n would still likely be biased, even when assuming that changes in firm scale on the right-hand side were perfectly exogenous. To see this, imagine we helicopter-dropped a random sales shock onto a firm that does not adjust either product quality or prices: even though product attributes stay unchanged, we would mechanically conclude that there are economies of scale in quality production ($\widehat{\beta_n} = \frac{1}{\bar{\sigma}_n - 1} > 0$). The reason is that demand shocks that one would usually want to exploit as IV for firm sales to estimate economies of scale in production, would in our setting, holding firm prices and quality constant, be mechanically interpreted as an increase in product quality.

To address this concern, we propose the following panel estimation strategy. Re-writing expression (3) for state-level demand instead of national-level, and again substituting for product quality from the optimal quality choice equation (12), we get:

$$\triangle \log(p_{nist}) = \beta_n \triangle \log(X_{nit}) - \frac{1}{N_z} \sum_z \frac{1}{\sigma_{nz}-1} \triangle \log\left(X_{nizst}\right) \qquad (21)$$

$$+\beta_n \triangle \log((\tilde{\gamma}_{nit}-\xi_n)/\mu_{nit}) + \eta_{nst} + \epsilon_{nist},$$

where subscript s indexes US states, η_{nst} are state-by-product module-by-half-year fixed effects, and $\triangle$ indicates a two-year change (four changes in our database starting from the first half year in 2006 until the end of 2014). As before, the second term on the right captures the demand-side relationship between sales and product unit values conditional on product quality, but this time at the state level. For instance, changes in firm productivity (and thus unit values on the left) conditional on product quality are captured by this term. The first and third terms capture the relationship between unit values and sales that is driven by changes in product quality. Following (12), firm changes in product quality are a function of national firm scale and the firm's composition of consumer taste parameters (where $\tilde{\sigma}_{nit}$ is part of μ_{nit}).

The advantage of writing the estimation equation in terms of state-level unit values on the left is that a helicopter drop of sales on a brand producer in another region of the US does not lead to a mechanical bias in β_n, unlike in the example above. The reason is that unless the firm changes its product quality in response, shocks to firm scale in other states have no effect on local unit values. The estimation also does not confound conventional economies of scale in production with economies of scale in product quality: if marginal costs fell with larger scale –holding quality constant–, this would be accounted for by the conventional demand relationship (second term on the right) discussed above. As above, we can re-write (21) for estimation as: $\triangle\log(\phi_{nist})=\beta_n\triangle\log(X_{nit}(\tilde{\gamma}_{nit}-\xi_n)/\mu_{nit})+\eta_{nst}+\epsilon_{nist}$.

The first remaining identification concern in (21) is correlated measurement errors between the left and right hand sides. A second concern is that firm changes in national sales are partly driven by taste shocks that could be correlated across states, which—holding constant product quality and unit values but not sales—would bias the estimate of β_n. To exploit plausibly exogenous variation in shocks to firm-level scale (21), we use leave-out mean changes in log firm sales across other states ($s' \neq s$), and using other product modules ($n' \neq n$). We then construct a weighted average of these leave-out mean changes in log firm sales using each firm's pre-existing share of total sales across different states.

This shift-share instrument for markup and composition-adjusted firm scale ($\triangle\log(X_{nit}(\tilde{\gamma}_{nit}-\xi_n)/\mu_{nit})$) is thus based on average changes in firm scale over time that exclude both the product group of the firm and the state in which $\triangle\log(\phi_{nist})$ on the left-hand side is observed. The identifying assumption of this strategy is that plausibly exogenous shocks to firm scale from other regions of the U.S. do not affect changes in state-level brand quality through other channels but firm scale.

5.3.3. Estimation results. We start in Table 2 by presenting reduced-form estimation results of the relationship between unit values or product quality on the left hand side and national firm sales on the right hand side. The raw empirical moment that is most directly informative of the degree of quality sorting across firm sizes is the fact that product unit values increase with national brand sales. This holds for both the cross-section of firms and for within-firm changes over time. It also holds in both OLS and IV estimations. In the cross-section, the IV addresses concerns about correlated measurement errors between unit values and firm scale and temporary taste shocks that could drive both left and right hand sides. In the panel data estimation, we have two-year changes in state-level log unit values on the left hand side, and we instrument the right hand side using plausibly exogenous changes in national firm sales (computed using the shift-share instrument described above). The IV point estimate of this panel IV specification in column 6 of Table 2 suggests that a 10% increase in a firm's national sales leads to a 0.7% increase in its unit value.

The same pattern of results holds when we replace unit values with our model-based measure of product quality on the left-hand side. In both the cross-section and the within-brand estimation product quality increases with national firm scale, and again this holds before and after addressing identification concerns using our instruments. In the panel IV specification in column 8, we find that a 10% increase in national firm sales leads to a 5.7% change in brand quality.

Table 3 proceeds to the estimation of β_n and ξ_n. To estimate these, the main difference to the previous reduced-form table lies in the additional inclusion of brand-level consumer compositions and firm markups on the right-hand side, as shown above in equations (20) and (21). The first panel reports the results when pooling all product groups, and the IV point estimates of the best-fitting parameter combination of β_n and ξ_n are not far from the reduced form results reported in Table 2. The second panel reports the technology parameter estimates separately for grocery and non-grocery product groups, and Table A.10 reports the estimation results separately for each product department. An interesting pattern emerges from the parameter estimates: in both the cross-sectional specification and the panel data approach, the IV point estimates of the economies of scale parameter in quality production are significantly larger for non-grocery product groups (*e.g.* health and beauty, merchandise) compared to grocery product groups. As indicated by the first stage F-statistics in the Supplementary Appendix table, the panel data estimation does not have sufficient power to precisely estimate β_n and ξ_n separately for each product department. For this reason, we use the precisely estimated parameters for grocery and non-grocery product groups reported in Table 3 for the counterfactual quantification in the following section. Following the discussion of identification concerns in the cross-section above, we use the IV panel estimation as our preferred parameter estimates, and report counterfactuals using the cross-section IV estimates as part of the robustness checks.

5.4. *Quantification of forces*

Armed with the preference and technology parameter estimates, we can check whether the calibrated model quantitatively replicates the main stylized fact documented in Figure 1. We can also use the calibrated model to quantify the forces underlying this observed relationship. There are multiple reasons for why the calibrated model may deviate from the observed estimates in Figure 1. The model imposes functional form assumptions on both demand and supply. It abstracts from other determinants of product quality across firms that operate independent of firm scale, and it does not capture temporary demand or supply shocks that are present in the observed sales data. In addition, the estimation of both the demand and supply-side parameters in the previous section is based on panel data (changes over time), whereas the moments we compare the calibrated model to are cross-sectional.

Following expressions (15) and (16), we can decompose the observed differences in weighted average firm sizes across consumption baskets into different forces on the demand and supply sides. In Figure 4, we depict different calibrated distributions of weighted average firm sizes across the aggregate consumption baskets of the five income groups alongside the observed moments in the data. We do this 18 times for each of the half-year periods in our dataset, and plot the mean outcomes for both the actual and calibrated moments.

In the first calibration, we only make use of the first part of expression (15) to predict the consumption choices of rich and poor income groups in a model world where the only source of heterogeneity between them is that they are subject to different estimated demand elasticities. That is, we predict the consumption shares of rich and poor income groups within product groups taking the quality and quality-adjusted prices of brands as given on the supply side in the data, assigning all households the same average taste-for-quality parameters $\overline{\gamma_n}$, but making use of the observed differences in their $\sigma_n(z)$ estimates. As depicted in Figure 4, household heterogeneity in price elasticities would, ceteris paribus, push poor households to consume from significantly larger firms compared to rich households—the opposite direction to what we observe in the data

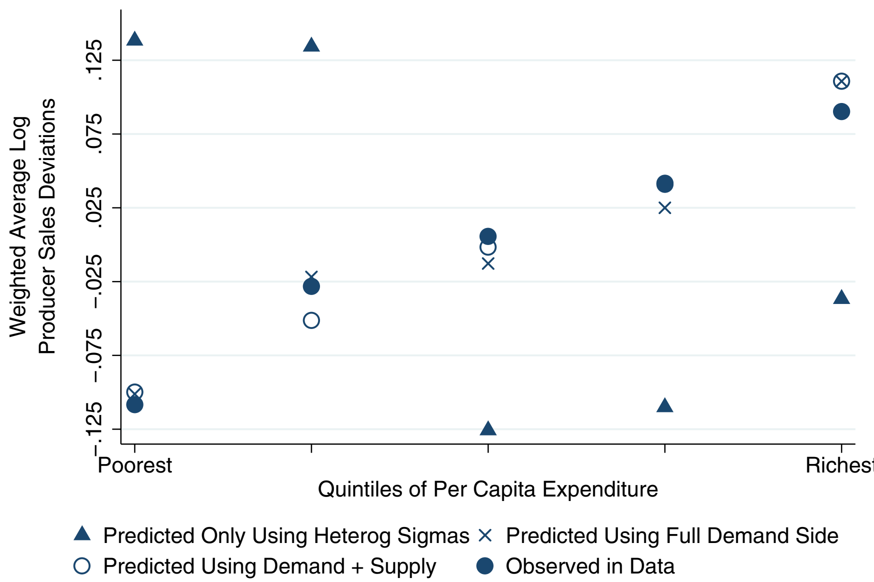


FIGURE 4

Decomposition of the underlying forces

Notes: The figure depicts predicted (model-based) and observed deviations in firm sizes across consumption baskets. See Section 5.4 for discussion.

in Figure 1 across individual households and in Figure 4 across the aggregate demand of different income groups in the data.

In the next calibration, we predict household consumption shares across the 5 income groups after also taking into account the second source of heterogeneity on the consumption side in expression (15): the fact that rich and poor households are estimated to value product quality differently. Again, we take as given the product quality and quality-adjusted prices on the supply side across brands in the data, and predict income group-specific consumption shares that are now taking into account both heterogeneity in σ_{nz} and in γ_{nz}. As shown in Figure 4, the fact that higher-income households are estimated to have significantly stronger tastes for product quality pushes in the opposite direction to the heterogeneity due to price elasticities, and dominates that first effect. The sum of the two effects in expression (15) closely replicates the differences in firm sizes across income quintiles documented in Figure 1.

In the final calibration, we fully endogenize both product choices on the consumer side and product choices on the firm side. That is, rather than predicting the consumption shares of income groups within product groups conditional on the available mix of product quality and quality-adjusted prices on offer across producers, we first predict the product quality choice across the firm size distribution using the equilibrium expression (16), and then let consumers optimally allocate budget shares on the demand side based on these predicted firm product choices. The only raw moments we use in these calibrations from the data is the observed distribution of firm sales across income groups for each of the 18 half years that we combine with the structure of the model and the estimated parameters to make predictions about the equilibrium differences in firm sizes across consumption baskets.

In addition to quantifying the (opposing) forces underlying the observed stylized fact in Figure 1, this exercise is useful to see to what extent the model that we use to solve for counterfactuals below is also able to rationalize the key sales and consumption patterns we observe in the baseline equilibrium. Reassuringly the calibrated model in 4 is able to closely replicate the observed differences in weighted-average firm sizes across the income distribution.

6. COUNTERFACTUALS

In this section, we use the calibrated model to quantify a new set of GE effects on household price indices in two policy counterfactuals. The first policy directly affects the distribution of disposable incomes on the demand side. The second policy directly affects profits across the firm size distribution on the supply side. In our framework, these policies can affect household price indices differently across the income distribution—both through direct effects and endogenous GE adjustments in firm scale, product quality choices, variable markups and exit/entry across the firm size distribution that affect rich and poor households differently. In the final section, we investigate the sensitivity of these counterfactual results across a number of alternative model and parameter assumptions. We also discuss the implications for the distribution of the gains from trade relative to a conventional Melitz (2003) model of heterogeneous firms.

6.1. *Counterfactual 1: progressive income tax reform*

One implication of the stylized fact in Figure 1 and the model we use to rationalize it in Section 4 is that changes to the distribution of disposable nominal incomes lead to GE effects on consumer price indices that tend to amplify the observed change in nominal inequality. To make this theoretical result concrete and put numbers on the mechanisms, we evaluate the implications of increasing the effective tax rate on the richest household group in our calibration (incomes above the 80th percentile) by 20 percentage points (from currently around 30–50%).

Through the lens of our model, this counterfactual allows us to relate to current policy debates in three main respects. First, this policy captures the historical change in US effective rates on incomes of the top 20% moving back to previous levels in the 1950s and 60s (Saez and Zucman, 2019).[41] Second, this policy closely corresponds to the counterfactual of moving the US from the current effective rate on the top 20% to the average effective rate on this group among Northern European countries.[42] Third, this policy change is also in line with the proposed tax reforms by two presidential candidates for the 2020 US elections (Sanders and Warren).[43]

The direct effect of this policy is a compression of the distribution of disposable nominal incomes. In line with the motivation behind this policy, we redistribute the tax revenues to the poorest 20% of U.S. households in our baseline counterfactual. Alternatively, we report results without redistributing the tax revenues to household disposable incomes. We solve for counterfactual changes 18 times, based on the observed brand sales to the five income groups for

41. While effective rates vary at a more granular level than the income quintiles we calibrate the model to, the counterfactual captures the historical change in effective rates on the sum of incomes above the 80th percentile (the income-weighted average change). Given both tax changes and taste for quality increase with incomes within this group, the results provide conservative estimates in this respect.

42. The average effective rate on incomes above the 80th percentile reported in the Luxembourg Income Study for Austria, Denmark, Finland, Germany, the Netherlands, Norway and Sweden is 52%.

43. See the tax sections of their campaign webpages (https://berniesanders.com, https://elizabethwarren.com) and effective tax calculator of their proposals at TaxJusticeNow.org. The Sanders proposal is close to the 20 percentage point increase above. The Warren proposal is closer to a 15 percentage point increase. We do not count health reform as part of taxation here, which would reinforce the progressivity.

each half-year cross-section in the scanner data and our estimates for the parameters $\sigma_n(z)$, $\gamma_n(z)$, β_n and ξ_n.[44]

We solve for the counterfactual equilibrium as described in Section 4.2.2, and decompose the mechanisms as derived in Supplementary Appendix 5. To compute confidence intervals that account for both sampling variation in the sales data across the 18 cross-sections and in the parameter estimates, we bootstrap the quantification 200 times for each half year of data. In each bootstrap, we draw the parameters $\sigma_n(z)$, $\gamma_n(z)$, β_n, and ξ_n from a normal distribution with a mean equal to the point estimate and a standard deviation equal to the standard error of the estimate.[45]

Figure 5 presents the counterfactual results. The left panel depicts the difference in the retail price index effect of the policy reform across the five U.S. income groups. As a result of the GE forces, we find that progressive income taxes gives rise to a meaningful amplification of the policy's direct effect, increasing its progressivity through price index effects. In particular, the bottom income quintile experience a 3 percentage point lower cost of living increase for retail consumption compared to the richest quintile (point estimate 2.94 with bootstrapped standard error of 0.99).

We can also decompose the inflation difference between the richest and the poorest quintiles, as outlined in Supplementary Appendix 5.B (Table A.11). The main drivers are related to both average and differential changes in firm scale and product quality across the distribution of initial firm sizes. As the policy leads to a reduction in the share of total sales to richer households in the economy, who value product quality the most, firms on average have incentives to downgrade product quality so that initial sales-weighted average product quality decreases. Since rich households value quality more, this average effect increases retail inflation among the rich relative to lower incomes, accounting for about 10% of the total difference in inflation.

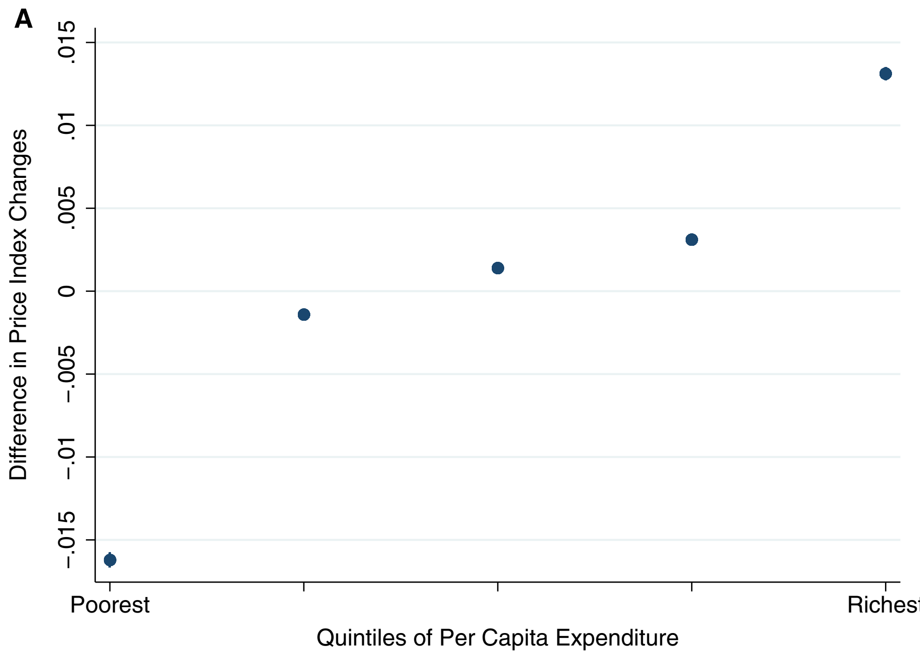
The main effect is that initial sales to rich households are particularly concentrated among initially large firms producing at higher product quality. The tax reform thus also leads to a compression of the firm size distribution. With economies of scale in quality production ($\beta_n > 0$), this translates into asymmetric effects on quality and quality-adjusted prices across the consumption basektes of rich and poor households. The right panel of 5 depicts this asymmetric effect on quality downgrading across the initial firm size distribution within product groups (with positive changes indicating reductions in product quality on the y-axis). On average, firms above the median size within product groups downgrade their product quality and vice-versa for firms below the median size. This effect accounts for about 80% of the overall inflation differential.

There is also a smaller effect operating through firm exit and love of variety. Since richer households have slightly lower estimated elasticities of substitution, firm exit increases their retail price index relative to poorer consumers (accounting for 6% of the overall difference).[46] We also quantify the implications of the same policy change, but without redistribution of the tax proceeds to household disposable incomes (Figure A.24). Reassuringly, we find that the effect on household inflation differences is quite similar, with about a 2.5 percentage point inflation differential between the richest and poorest quintiles. In Section 6.3 below, we find similar or

44. For each of the 18 half-year periods, we verify that the moments in the data and estimated parameter values satisfy the uniqueness conditions discussed in Section 4.2, Appendices 2 and 5. Our fitted model cannot perfectly match sales and quality valuations for each brand and each quintile. We can, however, perfectly fit the data by adding a multiplicative adjustment term ϵ_{niz} to quality valuations specific to each income quintile in expression (4). We then obtain the exact same counterfactual equations for *changes* in quality, sales and price indices as long as we hold these adjustment terms constant. Section 6.3 and Figure A.22 provide counterfactual results using different initial sales and consumption patterns to what we observe in the baseline equilibrium.

45. This is a parametric bootstrap (Horowitz, 2001). See *e.g.* Atkin *et al.* (2018) for a recent application.

46. Figure A.23 breaks up the price index effects across the different departments in the scanner data.



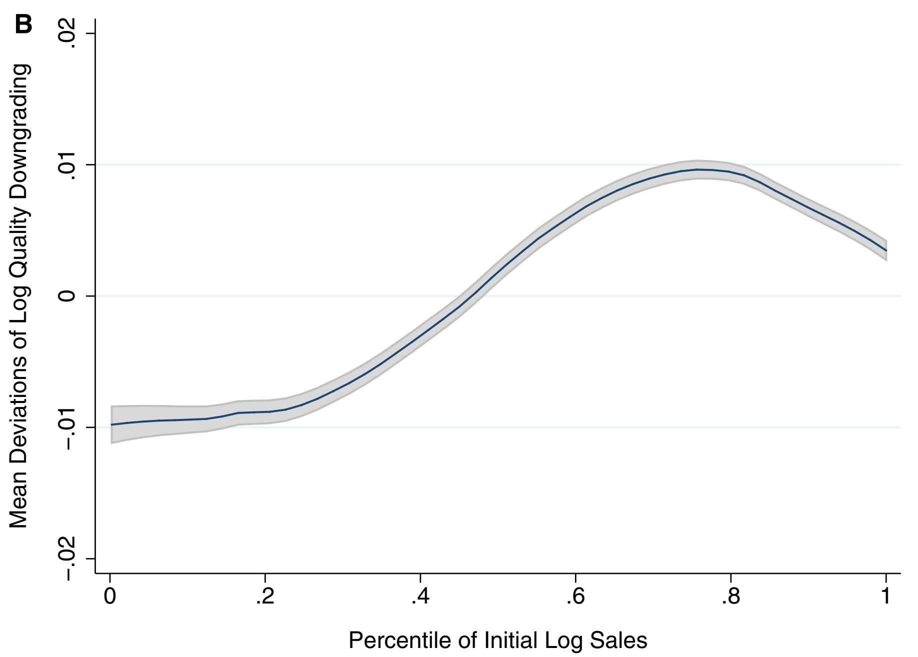


FIGURE 5

Counterfactual 1: inflation differences and quality downgrading due to more progressive income taxes

Notes: Both graphs display counterfactual changes averaged across 18 half-year cross-sections of data. 95% confidence intervals are based on robust standard errors. Panel A displays counterfactual inflation differences for retail consumption across income groups. Panel B displays deviations of output quality downgrading across percentiles of initial firm sales within product-module-by-period cells. See Section 6 for discussion.

slightly stronger results when revisiting the counterfactual across a number of alternative model or parameter assumptions.

6.2. *Counterfactual 2: closing loopholes in corporate taxation*

A second implication of our framework is that regulations that affect the profits of large and small firms differently give rise to differential price index effects (even in absence of GE adjustments), as these firms enter consumption baskets across the income distribution asymmetrically. To make this theoretical result concrete and put numbers on the mechanisms, we relate to the ongoing debate about closing loopholes in corporate taxation. A growing literature in public finance has documented larger possibilities for tax evasion among large U.S. corporations (*e.g.* Guvenen *et al.*, 2017; Wright and Zucman, 2018). A study by Bao and Romeo (2013) documents that the relationship between firm sales and effective corporate tax rates in the U.S.—which is monotonically and smoothly increasing up to the 95th percentile of firm sales—shows a sharp kink with a switch in the sign of the slope for the largest 5% of producers.

We use these findings to evaluate the impact of eliminating the kink in the otherwise smooth relationship between effective tax rates and firm size.[47] This policy change would lead to an increase of on average 5% in corporate taxes paid by the largest 5% of producers. This increase ranges from on average 1% at the 95th percentile to 11% at the 99th percentile of firm sizes. In our baseline counterfactual we follow the definition of firms as brands from Section 3 to define tax entities. Alternatively, we report results after using holding companies as the unit of taxation.

Figure 6 presents the counterfactual results. The left panel depicts the difference in the retail price index effect of the policy across the five U.S. income groups. We find that even this relatively modest adjustment in corporate taxation leads to a meaningful GE effect on differences in price indices between rich and poor households. This is in the order of a 1.5 percentage point lower inflation for retail consumption among the bottom 20% of US households compared to the top 20% (point estimate of 1.46, bootstrapped standard error of 0.34).

As above, we can decompose this differential impact (Table A.11). The first channel is the direct incidence of the policy, holding fixed initial product choices by firms and consumers. Since the largest firms sort into producing higher product quality, they represent a larger fraction of retail consumption sourced by the richest quintile compared to the poorest. These firms also experience higher tax increases that are passed on to consumer prices. This direct effect accounts for about one third (36%) of the overall difference in consumer inflation.

The second channel is again related to changes in firm scale and product quality across the firm size distribution. The largest firms are faced with a decrease in effective revenues and experience a reduction in their sales relative to firms not subject to the tax increase. To document this compression of the firm size distribution, we plot counterfactual changes in log firm sales as a function of initial percentiles of the firm size distribution (Figure A.25). Given economies of scale in quality production, and the large share of sales that the right tail of the size distribution represents, this leads to a reduction in the sales-weighted average product quality. The resulting changes in quality and quality-adjusted prices for consumption increases inflation relatively more among richer households, accounting for another 37% of the overall difference in price indices. As in the previous counterfactual, slight differences in love of variety imply that firm exit increases the retail price index more among richer households (accounting for 17% of the overall difference).[48] Finally, in the context of corporate taxation, tax entities may be holding companies rather than

47. To do so, we use estimates documented in Table 2 of Bao and Romeo (2013).

48. Figure A.26 breaks up the price index effect across the different departments in the scanner data.

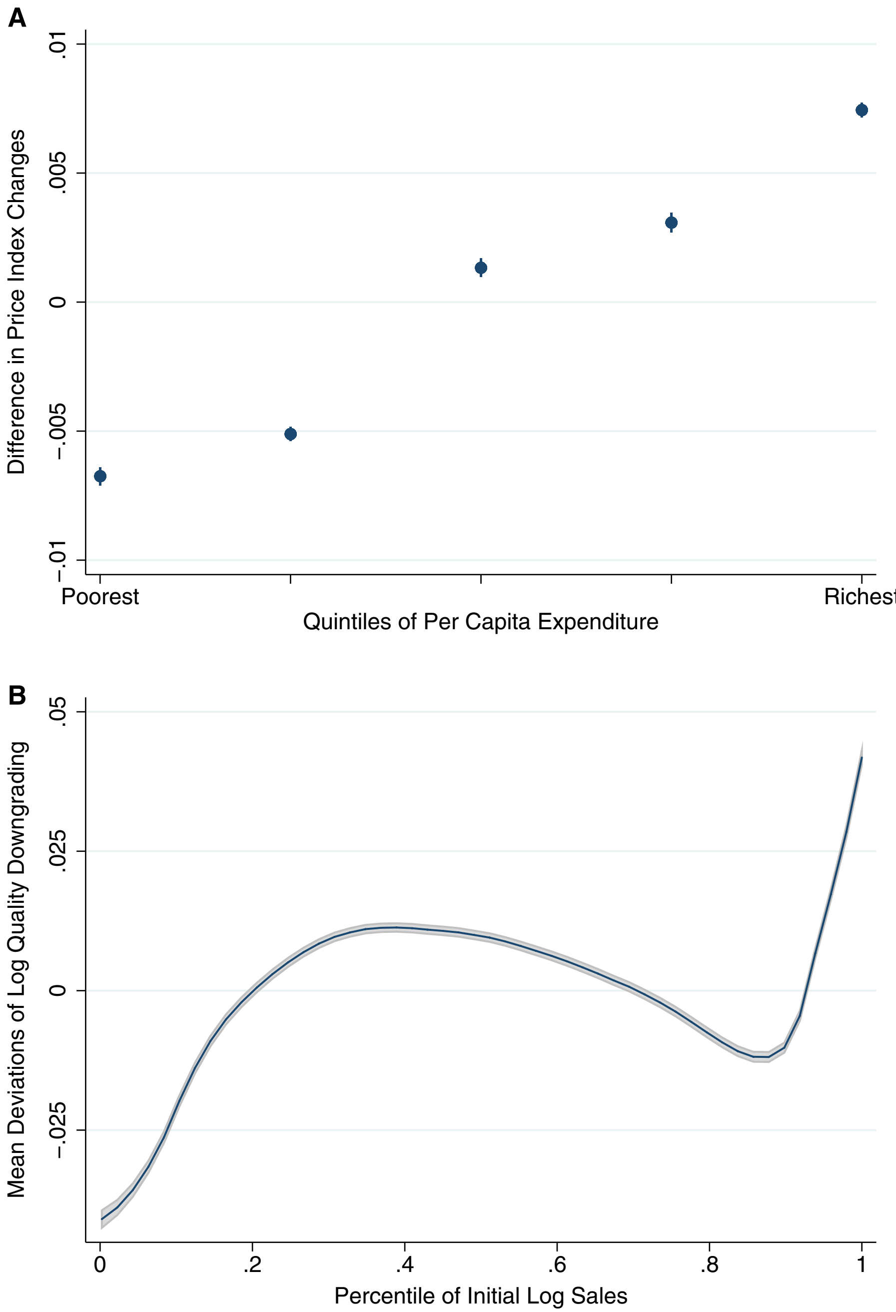


FIGURE 6

Counterfactual 2: inflation differences and quality downgrading due to closing loopholes in corporate taxation

Notes: Both graphs display counterfactual changes averaged across 18 half-year cross-sections of data. 95% confidence intervals are based on robust standard errors. Panel A displays counterfactual inflation differences for retail consumption across income groups. Panel B displays deviations of output quality downgrading across percentiles of initial firm sales within product-module-by-period cells. See Section 6 for discussion.

establishments producing brands. Reassuringly, we find that the counterfactual results are very similar (slightly stronger at 1.8 percentage points) when using holding companies instead of brands in the data to assign counterfactual tax changes (Figure A.27).

6.3. *Robustness and additional results*

In the final section, we explore the sensitivity of these findings to alternative model and parameter assumptions. First, as discussed in Section 5, we use alternative values for the demand elasticities and technology parameters compared to our baseline estimates for $\sigma_n(z)$, β_n, and ξ_n. Second, we re-estimate counterfactual outcomes after allowing household types (z) to endogenously change as a result of the counterfactual shocks. Third, we report counterfactuals under the more common assumption on market structure of monopolistic competition. Finally, we investigate the implications for the distribution of the gains from trade relative to a conventional Melitz (2003) model with two symmetric countries.[49]

6.3.1. Alternative parameter values. In line with the empirical literature in Industrial Organization and Quantitative Marketing, we find somewhat lower sales-to-price elasticities compared to the recent trade literature. To explore the sensitivity of our counterfactuals, we thus re-estimate each of them after multiplying our preferred estimates in Table 1 by a factor of 1.5 or 2 (keeping relative elasticities unchanged between income groups). Reassuringly, we find slightly larger effects compared to our baseline counterfactuals in all cases (Table A.12). Instead of using separate $\sigma_n(z)$ estimates by income group, we can also use the parametric specification in Section 5.1 as a function of income percentiles, shown in Panel C of Table A.8. This yields more continuous variation in $\sigma_n(z)$ across income groups compared to the baseline estimation for below and above median income groups. We find very similar counterfactual results as in our baseline specification (3.1 versus 2.9 and 1.4 versus 1.5). On the production side, we also report results after using the cross-sectional technology parameter estimates, instead of our preferred panel IV estimates in Section 5.3. We find slightly larger effects for both counterfactuals (3 versus 2.9 and 2.2 versus 1.5).

6.3.2. Endogenous changes in outside good consumption (z). As discussed in Section 4, we hold household income types z fixed in our baseline counterfactuals, ignoring that changes in real incomes may push some households across z group boundaries, and thereby affecting demand parameters $\alpha_n(z)$, $\sigma_n(z)$, and $\gamma_n(z)$ (see envelope theorem result in Supplementary Appendix 3). While appealing for not having to take a stance on the upper-tier utility function (1), this shuts down potential second-order effects on price indices due to changes in these parameters. In our theoretical framework, the change in consumer type is captured by changes in the consumption of the outside good (z). In Supplementary Appendix 3, we specify the upper-tier between retail consumption and the outside good, and quantify the change in outside good consumption induced by price index changes in the two counterfactuals above. Reassuringly, we find that the second-order adjustment channel has a negligible effect on outside good consumption ($< 0.3\%$), and thus a negligible effect on the counterfactual results. This may not be surprising, as the price index differentials we quantify above are in the order of

49. We also compute counterfactuals after assuming that, due to exogenous taste shocks, the initial equilibrium is subject to either more or less pronounced sorting of large firms selling more of their output to richer households than in the data (Figure A.22). Holding all model parameters as before, inflation differences between the rich and poor are amplified if initial firm sorting is stronger, and they are attenuated vice versa. See figure notes for additional details.

a 3 percentage point difference for retail consumption between rich and poor households. The fact that such a shock does not give rise to meaningful second-order effects through changes in household taste-for-quality or price elasticities—through pushing some households across quintiles—appears reasonable.

6.3.3. Monopolistic competition. Our model allows for large firms to take into account their market power through oligopolistic competition. A more common case with heterogeneous firms, following Melitz (2003), features a continuum of firms interacting under monopolistic competition. Even in this case, our framework with two-sided heterogeneity allows for variable markups across firms as a function of differences in their composition of consumer types. For completeness, we compute the two counterfactual equilibria in a model extension with monopolistic competition instead. Supplementary Appendix 5.A derives the system of counterfactual equations and Figure A.28 presents the counterfactual results. The point estimates are similar compared to the baseline results (allowing for oligopoly), and slightly stronger in the first counterfactual (3.5 versus 3.1 and 1.5 versus 1.5, respectively).

6.3.4. Implications for the distribution of the gains from trade. Finally, we investigate the implications for the distribution of the gains from trade. To do so, we introduce quality choice under two-sided heterogeneity into an otherwise standard Melitz model with monopolistic competition and two symmetric countries, and calibrate the model to the US data we use above (see Supplementary Appendix 5.C).[50] As in Melitz (2003), a decrease in trade costs induces a reallocation in which the largest firms expand through trade while less productive firms either shrink or exit. In our framework, better access to imported varieties and exit of domestic producers affects the price indices of rich and poor households asymmetrically. In addition, lower trade costs lead to heterogeneous changes in product quality and markups across firms. As a result, we find that a 10 percentage point bilateral increase in import penetration leads to a 3.5 percentage point lower retail price inflation for the richest 20% of households compared to the poorest 20% (Figure A.29). Relative to the conventional case—where heterogeneous firms are evenly represented across consumption baskets—we find that the stylized fact in Figure 1 and model we use to rationalize it imply a more unequal distribution of the gains from trade.

7. CONCLUSION

This article presents evidence that the widely documented presence of Melitz-type firm heterogeneity within sectors translates asymmetrically into the consumption baskets of households across the income distribution. To do so, we bring to bear detailed home and store scanner microdata that allow us to trace the national firm size distribution into the consumption baskets of individual households. We use the data to explore the underlying forces, and propose a quantitative GE model of quality choice under two-sided heterogeneity to rationalize the observed moments and explore policy counterfactuals.

We document significant differences in the weighted average firm sizes that rich and poor households source their consumption from. After quantifying a set of opposing forces, we find that this pattern is mainly explained by two features of household preferences and firm technology. On

50. We calibrate fixed trade costs such that half of of output is produced by exporting firms. We calibrate variable trade costs such that export sales of exporters equal 20% of their output. Combining these two statistics, about 10% of aggregate output is traded. The counterfactual is to reduce variable trade costs from an equilibrium with no trade to this new trade equilibrium.

the demand side, rich and poor households on average agree on their ranking of product quality evaluations within product groups. However, richer households value higher quality attributes significantly more compared to poorer households. On the production side, we estimate that producing higher output quality increases both the marginal and the fixed costs of production. Combined, these forces give rise to the endogenous sorting of larger, more productive firms into products that are valued relatively more by richer households.

These results have implications for the effect of policies and economic shocks on real income inequality. We find that the direct effect of progressive income taxation on inequality is amplified through asymmetric GE effects on household price indices, and that business regulations or trade liberalization that affect large and small firms differently give rise to new distributional implications. Underlying these findings is a rich interplay of firm adjustments in scale, product quality, variable markups, exit and entry that vary across the firm size distribution and thus affect rich and poor households differently.

Overall, our findings suggest that firm heterogeneity affects real income inequality in more complex ways than through the nominal earnings of workers, which have been the focus of the existing literature. These findings arise after introducing a basic set of features that we observe in the microdata—allowing for choice in product attributes by heterogeneous firms and households with non-homothetic preferences—into an otherwise standard economic environment. Empirically, these findings emphasize the importance of capturing changes in price indices at a granular level of product aggregation for both the measurement of changes in real income inequality over time and for studying the effects of policies and economic shocks.

Acknowledgments. We would like to thank Mary Amiti, Costas Arkolakis, David Atkin, Kirill Borusyak, Pablo Fajgelbaum, Cecile Gaubert, Jessie Handbury, James Harrigan, Xavier Jaravel, Thierry Mayer, Volker Nocke, Steve Redding, Andres Rodriguez-Clare, Nicolas Schultz, Eric Verhoogen, Jonathan Vogel, and participants at multiple seminars and conferences for helpful comments and discussions. Dmitri Koustas and May-Lyn Cheah provided outstanding research assistance. Researcher(s) own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online. And the replication packages are available at https://dx.doi.org/10.52xx/zenodo.4920923.

Data Availability Statement

The data underlying this article were provided by the Nielsen Company (US), LLC through the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. Details on how to obtain access and supplementary material for this article are available at https://dx.doi.org/10.5281/zenodo.4920923.

REFERENCES

AGUIAR, M. and HURST, E. (2007), “Measuring Trends in Leisure: The Allocation of Time over Five Decades”, *The Quarterly Journal of Economics*, **122**, 969–1006.

ARGENTE, D. and LEE, M. (2021), “Cost of Living Inequality During the Great Recession” *Journal of the European Economic Association*, **19**, 913–952.

ARKOLAKIS, C., COSTINOT, A., DONALDSON, D., et al. (2018), “The Elusive Pro-competitive Effects of Trade”, *The Review of Economic Studies*, **86**, 46–80.

ATKIN, D., FABER, B. and GONZALEZ-NAVARRO, M. (2018), “Retail Globalization and Household Welfare: Evidence from Mexico”, *Journal of Political Economy*, **126**, 1–73.

BALDWIN, R. and HARRIGAN, J. (2011), “Zeros, Quality, and Space: Trade Theory and Trade Evidence”, *American Economic Journal: Microeconomics*, **3**, 60–88.

BAO, D.-H. and ROMEO, G. C. (2013), "Tax Avoidance and Corporations in the United States - The Effective Tax Rate Abnormality for the Top Five Percent by Corporate Size", *Journal of Applied Business and Economics*, **14**, 88–100.

BARTELSMAN, E., HALTIWANGER, J. and SCARPETTA, S. (2013), "Cross-country Differences in Productivity: The Role of Allocation and Selection", *The American Economic Review*, **103**, 305–334.

BASTOS, P., SILVA, J. and VERHOOGEN, E. (2018), "Export Destinations and Input Prices", *American Economic Review*, **108**, 353-392.

BERNARD, A. B., JENSEN, J. B., REDDING, S. J., et al. (2007), "Firms in International Trade", *The Journal of Economic Perspectives*, **21**, 105–130.

BIJMOLT, T. H., VAN HEERDE, H. J. and PIETERS, R. G. (2005), "New Empirical Generalizations on the Determinants of Price Elasticity", *Journal of Marketing Research*, **42**, 141–156.

BLOOM, N. and VAN REENEN, J. (2007), "Measuring and Explaining Management Practices across Firms and Countries", *The Quarterly Journal of Economics*, **122**, 1351–1408.

BRODA, C. and WEINSTEIN, D. (2010), "Product Creation and Destruction: Evidence and Price Implications", *American Economic Review*, **100**, 691–723.

BURSTEIN, A. and VOGEL, J. (2017), "International Trade, Technology, and the Skill Premium", *Journal of Political Economy*, **125**, 1356–1412.

CARD, D., HEINING, J. and KLINE, P. (2013), "Workplace Heterogeneity and the Rise of West German Wage Inequality", *The Quarterly Journal of Economics*, **128**, 967–1015.

DAVIS, D. R. and HARRIGAN, J. (2011), "Good Jobs, Bad Jobs, and Trade Liberalization", *Journal of International Economics*, **84**, 26–36.

DEKLE, R., EATON, J. and KORTUM, S. (2007), "Unbalanced Trade", *The American Economic Review*, **97**, 351–355.

DELLAVIGNA, S. and GENTZKOW, M. (2019), "Uniform Pricing in US Retail Chains", *The Quarterly Journal of Economics*, **134**, 2011–2084.

DINGEL, J. I. (2016). "The Determinants of Quality Specialization", *The Review of Economic Studies*, **84**, 1551–1582.

FAJGELBAUM, P., GROSSMAN, G. and HELPMAN, E. (2011). "Income Distribution, Product Quality, and International Trade", *Journal of Political Economy*, **119**, 721–765.

FALLY, T. (2018), "Integrability and Generalized Separability" (NBER Working Paper 25025).

FEENSTRA, R. C. (1994), "New Product Varieties and the Measurement of International Prices", The American Economic Review, **84**, 157–177.

FEENSTRA, R. C. and ROMALIS, J. (2014), "International Prices and Endogenous Quality", *The Quarterly Journal of Economics*, **129**, 477–527.

FIELER, A. C., ESLAVA, M. and XU, D. Y. (2018), "Trade, Quality Upgrading, and Input Linkages: Theory and Evidence from Colombia", *American Economic Review*, **108**, 109–46.

FRIAS, J. A., KAPLAN, D. S. and VERHOOGEN, E. A. (2009), "Exports and Wage Premia: Evidence from Mexican Employer-Employee Data" (Working Paper, Columbia University).

GUVENEN, F., MATALONI Jr, R. J., RASSIER, D. G., et al. (2017), "Offshore Profit Shifting and Domestic Productivity Measurement" (NBER Working Paper 23324).

HALLAK, J. C. and SIVADASAN, J. (2013), "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia", *Journal of International Economics*, **91**, 53–67.

HANDBURY, J. (2019), "Are Poor Cities Cheap for Everyone? Non-homotheticity and the Cost of Living across US Cities" (NBER Working Paper 26574).

HANDBURY, J. and WEINSTEIN, D. E. (2014), "Goods Prices and Availability in Cities", *The Review of Economic Studies*, **82**, 258–296.

HAUSMAN, J. (1999), "Cellular Telephone, New Products, and the CPI", *Journal of Business & Economic Statistics*, **17**, 188–194.

HAUSMAN, J. and LEIBTAG, E. (2007), "Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart", *Journal of Applied Econometrics*, **22**, 1157–1177.

HELPMAN, E., ITSKHOKI, O., MUENDLER, M.-A., et al. (2017), "Trade and Inequality: From Theory to Estimation", *The Review of Economic Studies*, **84**, 357–405.

HOROWITZ, J. L. (2001), "The Bootstrap", in Heckman, J. J. and Leamer, E. (eds), *Handbook of Econometrics* (Elsevier), Vol. 5 3159–3228.

HOTTMAN, C. J., REDDING, S. J. and WEINSTEIN, D. E. (2016), "Quantifying the Sources of Firm Heterogeneity", *The Quarterly Journal of Economics*, **131**, 1291–1364.

HSIEH, C.-T. and KLENOW, P. J. (2009), "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics*, **124**, 1403–1448.

JARAVEL, X. (2019), "The Unequal Gains from Product Innovations: Evidence from the US Retail Sector", *The Quarterly Journal of Economics*, **134**, 715–783.

JOHNSON, R. (2012), "Trade and Prices with Heterogeneous Firms", *Journal of International Economics*, **86**, 43–56.

KELLER, K. L., PARAMESWARAN, M. and JACOB, I. (2011), Strategic Brand Management: Building, Measuring, and Managing Brand Equity (Prentice Hall).

KUGLER, M. and VERHOOGEN, E. (2012), "Prices, Plant Size, and Product Quality", *The Review of Economic Studies*, **79**, 307–339.

MCFADDEN, D. and TRAIN, K. (2000), "Mixed MNL Models for Discrete Response", *Journal of Applied Econometrics*, **15**, 447–470.

MELITZ, M. (2003), "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity", *Econometrica*, **71**, 1695–1725.
NEVO, A. (2000), "Mergers with Differentiated Products: The Case of the Ready-to-eat Cereal Industry", *The RAND Journal of Economics*, **31**, 395–421.
NIELSEN COMPANY. (2016a). NielsenIQ: Consumer Panel Dataset. https://www.chicagobooth.edu/research/kilts/datasets/nielseniq-nielsen. (Last accessed for download: 03/11/2016).
NIELSEN COMPANY. (2016b), NielsenIQ: Retail Scanner Dataset. https://www.chicagobooth.edu/research/kilts/datasets/nielseniq-nielsen (last accessed for download: 03/11/2016).
PETERS, M. (2020), "Heterogeneous Mark-ups, Growth and Endogenous Misallocation", *Econometrica*, **88**, 2037–2073.
RAY, W. (2019), "Cyclicality of Markups: Evidence from Heteroskedasticity-identified Estimation" (Working Paper, London School of Economics).
REDDING, S. J. and WEINSTEIN, D. E. (2020), "A Unified Approach to Estimating Demand and Welfare Changes", *The Quarterly Journal of Economics*, **135**, 503–560.
SAEZ, E. and ZUCMAN, G. (2019), The Triumph of Injustice: How the Rich Dodge Taxes and How to Make Them Pay (New York: WW Norton & Co).
SAMPSON, T. (2014), "Selection into Trade and Wage Inequality", *American Economic Journal: Microeconomics*, **6**, 157–202.
SODERBERY, A. (2015), "Estimating Import Supply and Demand Elasticities: Analysis and Implications", *Journal of International Economics*, **96**, 1–17.
SONG, J., PRICE, D. J., GUVENEN, F., et al. (2018), "Firming Up Inequality", *The Quarterly Journal of Economics*, **134**, 1–50.
SUBRAMANIAN, S. and DEATON, A. (1996), "The Demand for Food and Calories", *Journal of Political Economy*, **104**, 133–162.
SUTTON, J. (1998), *Technology and Market Structure: Theory and History* (The MIT Press).
WRIGHT, T. and ZUCMAN, G. (2018), "The Exorbitant Tax Privilege" (NBER Working Paper 24983).